

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:21:40 ; Search time 127 Seconds

(without alignments)
369.954 Million cell updates/sec

Title: US-09-917-384-4

Perfect score: 2249
Sequence: 1 ATHVDPYAGATFEVNPYWA.....ACQWPPAQFDOLVANARPAV 423

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1023	45.5	1010	19	AAW34989	Teredinibacter end
2	999	44.4	423	17	AAW90715	Thermostable cellu
3	999	44.4	596	17	AAW90716	Thermostable cellu
4	569	25.3	395	21	AAW84797	Amino acid sequenc
5	562.5	25.0	476	19	AAW44853	Humicola insolens
6	562.5	25.0	476	19	AAW44827	Humicola insolens
7	558.5	24.8	473	20	AAW01077	H. insolens cellula
8	546	24.3	471	6	AAW50308	Cellobiohydrolase
9	538	23.9	471	16	AAW77262	T. longibrachiatum
10	538	23.9	471	17	AAW02025	Trichoderma cellulo
11	534	23.7	457	18	AAW25789	Acromonium cellulo

12	449.5	20.0	446	20	AAV01076	H. insolens Cel6B
13	439.5	19.5	449	19	AAW56739	Orpinomyces celluli
14	431.5	19.2	432	21	AAW18414	Amino acid sequenc
15	411	18.3	459	19	AAW56738	Orpinomyces celluli
16	244.5	10.9	449	18	AAW18210	Cellobionomyces flumi
17	231	10.3	490	20	AAW95602	Saccharothrix aust
18	181.5	8.1	321	22	AAW70839	S. halstedii cellu
19	120.5	5.4	1790	22	AAW71739	Drosophila melanog
20	108.5	4.8	1095	10	AAW97052	Sequence encoded b
21	107	4.8	639	22	AAW12935	Dactylium dendroid
22	107	4.8	639	22	AAW12935	Dactylium dendroid
23	107	4.8	639	22	AAW12939	Dactylium dendroid
24	107	4.8	639	22	AAW12941	Dactylium dendroid
25	107	4.8	680	22	AAW63582	Amino acid sequenc
26	107	4.8	681	22	AAW63583	Synthetic amino ac
27	106	4.7	639	22	AAW12964	Dactylium dendroid
28	105.5	4.7	639	22	AAW12962	Dactylium dendroid
29	105	4.7	632	16	AAW77558	p45 metallopeptase
30	105	4.7	639	22	AAW12953	Dactylium dendroid
31	104	4.6	7068	22	AAW12960	Dactylium dendroid
32	104	4.6	639	22	AAW12940	Dactylium dendroid
33	104	4.6	639	22	AAW12946	Dactylium dendroid
34	103.5	4.6	1472	22	AAW10144	Dactylium dendroid
35	103	4.6	639	22	AAW59423	Dactylium dendroid
36	103	4.6	639	22	AAW12940	Dactylium dendroid
37	103	4.6	639	22	AAW12946	Dactylium dendroid
38	103	4.6	639	22	AAW12946	Dactylium dendroid
39	102	4.5	639	22	AAW12934	Dactylium dendroid
40	101.5	4.5	639	22	AAW12928	Dactylium dendroid
41	101	4.5	384	21	AAW64472	Arabidopsis thalia
42	101	4.5	421	21	AAW64471	Arabidopsis thalia
43	101	4.5	502	21	AAW64470	Novel human secret
44	100	4.4	1888	22	AAW31049	Fusarium oxysporum
45	99.5	4.4	388	17	AAW05845	

ALIGNMENTS

RESULT 1	AAW34989	standard; Protein: 1010 AA.
ID	AAW34989	
XX	AAW34989	
AC	AAW34989	
XX	21-MAY-1998	(first entry)
DE	Teredinibacter endoglucanase:	
XX	Endoglucanase: cellulase; carboxymethylcellulose; cellulose;	
KW	biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;	
KW	thermostable enzyme; thermophilic; glycosidase.	
XX	Teredinibacter sp. (Clone 426P1).	
OS	WO9744361-A1.	
XX	27-NOV-1997.	
PD	22-MAY-1997;	97WO-US08793.
XX	22-MAY-1997;	97WO-US08793.
PR	22-MAY-1996;	96US-0651572.
XX	(RECO-) RECOMBINANT BIOCATALYSIS INC.	
PA	Lam DE, Mathur EJ;	
XX	WPI: 1998-018435/02.	
XX	N-PSDB; AAT94197.	
DR	Endoglucanase(s), preferably form archaeal bacterium, AEPit 1a -	
PT	useful to degrade carboxymethylcellulose and hydrolyse of	
PT	beta-1,4-glycosidic bonds in cellulose	

BEST AVAILABLE COPY

PT Mutant Chrysosporium strain comprising nucleic acid encoding proteins,
 PT particularly industrially important enzymes such as cellulases,
 PT xylanases, pectinases, lipases and proteases -

Dislosure; Page 43-46; 92pp; English.

CC The present sequence represents a Chrysosporium C1 endoglucanase. The
 CC polynucleotide sequence comprises a promoter sequence, and can be
 CC linked to a nucleic acid sequence of the invention to control
 CC expression. The specification describes a mutant Chrysosporium
 CC strain which comprises a nucleic acid sequence encoding a polypeptide
 CC of interest, where the nucleic acid sequence is linked to an
 CC expression-regulating region and a secretion signal sequence.
 CC The mutant strain is useful for producing heterologous proteins
 CC and polypeptides of interest, e.g. industrially useful enzymes such
 CC as cellulases, xylanases, pectinases, lipases and proteases.

CC Sequence 395 AA;

Query Match 25.3%; Score 569; DB 21; Length 395;
 Best Local Similarity 35.5%; Pred. No. 3.7e-40;
 Matches 150; Conservative 54; Mismatches 148; Indels 70; Gaps 13;

QY 1 ATHVNDPVGATFFVNPVMAOEVSAAQNTNATLAAKRVSTYSTAVMDRIAINGV 60

DB 43 stuvfgy---tlhpnfryaeveaaaeaisdsalaekarkvadvgfllwltlenlg-- 97

QY 61 NGGPGITTYLDAALSOQOGTPEVIEIVYIDUPGRDCAALASNGELPATNAGLQTEYEQY 120

DB 98 -----rlpeal--edpcenivglviydlpyrdcaakaangel--kvgeldrykcy 145

QY 121 IPIPIASISNPKYSSIRIYTTIEPDSLPNATNMISIQACATVPEEGIEYALTKLHAI 180

DB 146 ldklaeilk--ahstafalviepdsipvlvnsdlqtcqgsaagyregvayalkqln-1 202

QY 181 PNVTYIMDAHSGMLGMPNNAAGVYOEYOKVKNASIGVNGIDGFPVTNTNPKPEPMT 240

DB 203 pnvmyyidagbhgywlgwdnlkprgaqelasyksaagspqrvgistlnvawg-----n 254

QY 241 ATQVGGQPVESANFYQWNPIDIDEADVADLYSRLVAAAGFPSSIGMLIDTLRNGMGPN 300

DB 255 awdggpge-fsdasdaqykcqneklyntfgaelksagmiph--aldtgnvtyc--- 308

QY 301 PPGSPATDVNFVNSKIDLRQHGKMGONAGLIGQPPQASPTDFPVAHLDAYWIRP 360

DB 309 -----lrdewgdwcunvgagfgvprltanbgd--eladafwvkr 345

QY 361 PGESDGTSAADPTGKKSDPPTVYTSYGVLTNNALPNSPIAGOWPFAOPDOLVANAR 420

DB 346 ggesdgtdsd-----aarydsfcgk-----pdafkpspeagtlwngaytemliknan 392

QY 421 PA 422

DB 393 ps 394

RESULT 5

ID AAM44853 standard; Protein; 476 AA.

AC AAM44853;

XX 31-JUL-1998 (first entry)

DE Humicola insolens cellulase NCE2 protein.

XX Humicola insolens; NCE1; NCE2; NCE4; cellulase; expression vector;

KM promoter; signal sequence; terminator; amylase; lipase; protease;

XX Humicola insolens.

OS

XX

PH Key Location/Qualifiers
 FT Peptide 1..23
 FT /label= signal
 FT Protein 23..476
 FT /label= Cellulase_NCE2

PN W09803667-A1.

PD 29-JAN-1998.

XX 24-JUL-1997; 97WO-JP02560.

XX 24-JUL-1996; 96JP-0195070.

XX (MEIJ) MEIJU SEIKA KAISHA LTD.

PA Aoyagi K, Hamaya T, Koga J, Kono T, Moriya T, Murakami T;

PI Murashima K, Sumida N, Watanabe M;

XX WPI; 1998-120786/11.

DR N-PSDB; AAV19377.

PT Mass production of proteins and peptides in Humicola species - using

PT expression vector containing the promoter, signal sequence and/or

PT terminator from the Humicola insolens NCE1 or NCE2 gene

PS Claim 8; Page 40-45; 63pp; Japanese.

CC The present sequence represents the Humicola insolens cellulase NCE2
 CC protein from the present invention. The present invention describes a
 CC method for the mass production of proteins and peptides in Humicola
 CC species, especially in Humicola insolens, using an expression vector
 CC which comprises the promoter, signal sequence and/or terminator
 CC regulatory sequences from the NCE1 or NCE2 gene of H. insolens. These
 CC are available in the plasmids pM3-1 (Escherichia coli JM109/pM3-1,
 CC FERM BP-5971) (for NCE1) and pM4-1 (E. coli JM109/pM4-1, FERM BP-5972)
 CC (for NCE2). The vector also contains a marker gene such as an antibiotic
 CC resistance gene (e.g. the tetracycline resistance gene from Streptomyces
 CC timentans). Proteins which can be expressed using this system include
 CC cellulase, amylase, lipase, protease, phytase and other enzymes.
 CC Specific expression vectors of the invention are pMKD01 (for Humicola
 CC NCE3 cellulase gene), pEGD01 (for Humicola NCE4 cellulase gene) and
 CC pHEM02 (for Humicola NCE4 cellulase gene). The expression system allows
 CC the efficient production of proteins and peptides in a Humicola host.
 CC Using the expression system high amounts of protein (>4.5 g/l) can be
 CC obtained.

SO Sequence 476 AA;

Query Match 25.0%; Score 562.5; DB 19; Length 476;
 Best Local Similarity 33.7%; Pred. No. 1.7e-39;
 Matches 147; Conservative 54; Mismatches 146; Indels 89; Gaps 15;

QY 1 ATHVNDPVGATFFVNPVMAOEVSAAQNTNATLAAKRVSTYSTAVMDRIAINGV 59

DB 113 aeyngnprlegvgqlwannyrsewhltpqldpalaasavaepsfqwidrvtyd- 171

QY 60 VNGGPGITTYLDAALSO-----QOGTPE-VIEIYIYIDUPGRDCAALASNGELPATNAGL 113

DB 172 -----tlivetsetraangagpnpyaaglyvyldpdcdaaasngewalannga 223

QY 114 QTEFOYIDPIASISNPKYSSIRIYTTIEPDSLPNATNMISIQACATVPEEGIEYALTKLHAI 173

DB 224 nnyk-gylirirellis--fsdvrtllvlepdalamvnmnvakscgaasyreltlyla 280

QY 174 LTRKHAIPNVYIYMDAASHGMLGMPNNAAGVYOEYOKVKNASIGVNGIDGFPVTNTNPKPEPMT 232

DB 281 lqyld-lphvanymdagbhgywlgwpanlqpaaelaklyedagkpravglatnvanyna 339

QY 233 -----PLKEPMTATQVGGQPVESANFYQWNPIDIDEADVADLYSRLVAAAGFPSSIG 285

DB 340 waispppytsp-----npydekhyiaefrplleagrfpag-- 376

Oy		286	MIDILRNQMGSPNEPTGPDSTADVTWTFVNSKXIDLROHRLCNCNONGACLGOPQASPT	345
Dd		377	fIvdggrsg-----kqptlg-----gkeghnchncalgtfgmtpant-	413
Oy		346	DEPNAMHLDAIVYIKRPGESDGTSGSASPDTGTKSDPMCDPFTYTTSYGVLTNALPSNDIAG	405
Dd		414	-_ghpyvdaftvvwvkpgsedcgt----dttaarydyhcgs-----ledalkpap eag	458
Oy		406	GMEPAOFODLVANARP	421
Dd		459	qwfygyfegllrnanp	474
SQ	Sequence	476 AA:		
ID	AAM44827 standard; Protein;	476 AA.		
AA	AAM44827			
XX	AAM44827;			
Dt	07-JUL-1998	(first entry)		
De	H. insolens cellulase NC2 protein sequence.			
KM	Cellulase; primer; PCR; amplification; probe; hybridisation; hydrolysis;			
Xx	cellulose; sugar; enzyme.			
OS	Humicola insolens.			
FH	Key	Location/Qualifiers		
FT	Peptide	1..23	/note= "signal peptide"	
FT	Protein	24..476	/note= "NC2 cellulase mature protein"	
PN	JF08126492-A.			
XX	21-MAY-1996.			
PF	02-NOV-1994;	94JP-0269293.		
PR	02-NOV-1994;	94JP-0269293.		
PA	(MEIJ) MEIJI SEIKA KAISHA LTD.			
DR	WPJ: 1998-056048/O6.			
DR	N-PsDB: AAV19281.			
Novel	cellulase component NCE2 - Isolated from Humicola insolens and used in detergents , as a feed additive etc.			
Discloure:	Page 10-12; 15pp; Japanese.			
This	is the amino acid sequence of the NC2 cellulase from Humicola insolens . The protein was isolated from cultured cells and the N-terminal peptide sequence determined. The protein was also cleaved with V8 protein to generate several peptide fragments. These were also sequenced. Degenerate primers were designed based on the sequences of the protease fragments (see AAV19282-V19283) and used to amplify a fragment of the gene for use as a probe to screen a genomic library. The remainder of the sequence was obtained and determined using primers AAV19284-V19299. The cellulase can be used to hydrolyse cellulose into sugar, and as an enzyme in industry, fibre-processing agent, feed additive, or digestive agent.			

Query Match	25.0%;	Score 562.5;	DB 19;	Length 476;
Best Local Similarity	33.7%;	Pred. No. 1.7e-39;		
Matches 147;	Conservative 54;	Mismatches 146;	Indels 89;	Gaps 15

```
OY      1 ATHNDNPAQCTFFPVNNVMAOEAOSEAAO--TNATLAAKKRRVSTYSTVAWMDRIIAAG   59
Db      113 asyngnpefgeqldaanmyysevhtlaipdltcpalraasaavaevpsfqwdrnvtvd-    171
OY      60 VNGGEGLTYYDAALSO-----QQGTPE-VIEIVTYDLPGRDCALAASNGELPATACL   113
Db      172 -----tlivetlsetiraanggangnpyaaqlvyvdlpdrtcdcaaaaengewalaanga   223
OY      114 QTEYOXYDIPASTLSNPKXSSLRIVYIIEPDSLPNAYTMMSIQACATVPVYEQGEIEFA   173
Db      224 mnyk-gynritrellis--fsdvrtllivsepbdsilamntlmnvakcsagaacyreltlya   280
OY      174 LTKLHAIPNVYIYUDDAASHGMLGMPNNAASGYVOEVOKVLNASIGVANGIDGFVTNTANTY-   232
Db      281 lkqidl-phvamymadgaehagwlgypnanlgraaelfakiyledagkravzglatunanya   339
OY      233 -----PLKRFMTATOOVGGGQVEBSANTYOMNPPRODDEADYAVDLXSLUYAAGFPSSSIG   285
Db      340 wslsspppytsr-----nnpydekhyleaftrpllergfpdq--   376
OY      286 MLILTLRNGWGSPNEPTGPSTATDVNTFVNQSKIDLROIHGKLGCNONGALCOPPOASPT   345
Db      377 flvdegatsg----kqrpgs-----qkewghbcmalgsgfgmiprant-   413
OY      346 DFNANHLDAYVIKRPESDSGTSAASDPPTTGKKSDDPCDFPTTYSIGVLTNALPNSPIAG   405
Db      414 --gbhydafwkvprkgsegdfts---dtlaarydyhcg-----ledalkpapeeg   458
OY      406 QMPPAFQDOTLANARP   421
Db      459 qwfqayfeqllrmannp   474
```

RESULT	7
AAV01077	
ID	AAV01077 standard; Protein; 473 AA.
XX	
AC	AAV01077;
XX	
DT	08-JUN-1999 (first entry)
XX	
DE	H. insolens Cel6A fungal cellulase protlen sequence.
XX	
KW	Cel6B; Cel6A; fungal cellulase; cleaning composition; conditioner;
KW	cellulolytically active protein; endo-1,4-beta-glucanase; enzyme;
KW	fabric softener; Humicola-like cellulase; glycosyl hydrolase family 6
KW	detergent composition.
XX	
OS	Humicola insolens.
XX	
PN	WO9901544-A1.
XX	
PD	14-JAN-1999.
XX	
PF	02-JUL-1998; 98WO-DK00299.
XX	
PR	04-JUL-1997; 97DK-0000813.
XX	
PA	(NOVO) NOVO-NORDISK AS.
XX	
PI	Andersen KV, Damgaard B, Lund H, Nielsen JB, Schultein M;
XX	
DR	WI: 1999-106046/09.
XX	
DR	N-PSDB; AAX27948, AAX27958.
XX	
PT	Cleaning composition containing Humicola endo-beta-1,4-glucanase
XX	
PT	useful as detergent compositions or additives, or as fabric
XX	
PS	conditioners
XX	
PS	Example 3; Page 243-245; 271pp; English.
XX	
CC	This sequence is the Humicola insolens Cel6A fungal cellulase.
CC	The invention relates to a cleaning composition (A) that contains at

XX	AA077262;	
AC		
XX	13-DEC-1995 (first entry)	
XX		
DE	T. longibrachiatum exo-cellobiohydrolase CBHII.	
XX		
KW	Cellulase; cellulose; signal; catalytic core; cellulase binding linker.	
XX		
OS	Trichoderma longibrachiatum.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..24
FT		/label= signal
FT		/note= "seq id no 28"
FT	Domain	25..46
FT		/label= cellulose binding domain
FT		/note= "seq id no. 4"
FT	Region	47..89
FT		/label= linker
FT		/note= "seq id no 20"
FT	Domain	90..471
FT		/label= catalytic core
FT		/note= "seq id no 12"
XX		
PN	MO9516782-A.	
XX		
PD	22-JUN-1995.	
XX		
PE	19-DEC-1994; 94MO-US14163.	
XX		
PR	17-DEC-1993; 93US-0169948.	
XX		
PA	(GEMV) GENENCOR INT INC.	
XX		
PI	Clarkson KA, Collier KD, Fowler T, Iarens E, Ward M;	
XX		
DR	WPI; 1995-231574/30.	
DR	N-PSDB; AA091282.	
XX		
PT	Pure, truncated fungal cellulase protein from Trichoderma - useful to	
PT	reduce or eliminate dye, colourant or pigment back-staining or	
PT	redeposition in stone-washing or bio-polishing	
XX		
PS	Claims 12, 39, 40, 41; Figure 2; 105pp; English.	
XX		
CC	Figure 2 depicts the genomic DNA and AA sequence of CBHII derived	
CC	from T. longibrachiatum. A truncated fugal cellulase with the AA	
CC	sequence in SEQ ID no. 4 is claimed. Also claimed are DNA gene	
CC	fragments encoding sequences: SEQ ID 12; SEQ ID 12 and 20; SEQ	
CC	ID 12, 20 and 16. Genes for CBHII have been isolated	
CC	from T. longibrachiatum and the protein domain structure has been	
CC	confirmed (Shoemaker, S. et al., 1983, Bio/Technology, 1, 691-696;	
CC	Teerl, T. et al. 1983, Bio/Technology 1, 696-699 and Teerl, T. et al.,	
CC	1987, Gene, 51, 43-52).	
XX		
SO	Sequence 471 AA;	
OY	Query Match 23.9%; Score 538; DB 16; Length 471;	
Db	Best Local Similarity 33.3%; Pred. No. 2.1e-37;	
	Matches 144; Conservative 51; Mismatches 155; Indels 82; Gaps 13;	
OY	1 ATHVNPAPGATFEFNPWAOGEVGEAANQNTATLAAKRVSTYSTAVAMD----- 52	
Db	110 atygsnpitvgyvpanayasevsalslpsltgamataaaavakvpsfwmldtlaktplm 169	
OY	53 --RIAAINGVNGGPELITTYLDALLSQOOGTPEVEIETIYDLPGRDCAALASNGELPATA 110	
Db	170 egtlaidittankng--nyag-----qfvaidipdtdcaalaasngeslad 213	
OY	111 AGIQTETQYIDPIASILSNPKYSSLRIYTIIEPDSLPAVATNMSIOACATVAPYEQGI 170	

Accession	Protein Name	Gene Name	Species	Length (aa)	Weight (kDa)	PI	Ref						
Db	214 g9vaykx-nyldtrqiv--veysdlrlfilylepdsla	nyldtrqiv	yeasts	121	13.5	4.10	1						
Qy	171 EFWALKRLAIPNPVYLYMDAAHSGWLGMPNNA	SGVVOEYQVKLVN	SGVNGIDGFWPTAN	230	23.0	4.10	1						
Db	271 nyavagqln-lpnvamyldaghnagwlgp	nyavagqln	nyavagqln	329	32.9	4.10	1						
Qy	231 YTPLEKPEPMTATQQVGQPVESANFYQMN	PDIDADYAVDL	YSLVAVGPPSSIGMLDT	290	29.0	4.10	1						
Db	330 yv-----gmnlteppaytfgnavaynekiy	hagipllanhgw-sna	ffltdq	375	37.5	4.10	1						
Qy	291 LRNGGCGNEPPTGSPETADVMTFVNVQSK	IDLRHRLGLCMQNGAGLQPPQASPTDEPNA	350	35.0	4.10	1							
Db	376 g9s9-----kqpgdg-----qqqwg	gdwcnvigtgfglrpsa	ntlgd--s	410	41.0	4.10	1						
Qy	351 HLDAYVMIKPGEESDGTSAASDPPTGKKSD	PDPTVYTSYGVLTNALPNSPIAGMFP	410	41.0	4.10	1							
Db	411 lldstvwkhpqsgcdgtssap-----rfdhc	-----alpdalqpagaqawfga	455	45.5	4.10	1							
Qy	411 QFDQLVANARPA	422											
Db	458 yfvqlltnaps	469											
RESULT 10													
AAW02025	standard; Protein: 471 AA.												
AAW02025													
AAW02025													
XX	28-OCT-1996	(first entry)											
XX	Trichoderma cellobiohydrolase II.												
XX	Trichoderma cellobiohydrolase II; CBHII; cellulase; cellulose; denim; stonewashing; dye redeposition; backstaining.												
XX	Trichoderma longibrachiatum.												
XX													
XX	Key	Location/Qualifiers											
XX	Peptide	1..24											
XX	Protein	/label= Sig_peptide											
XX	Domain	25..471											
XX	Region	/label= Mat_protein											
XX	Domain	25..63											
XX	Region	/label= Cellulose_binding_domain											
XX	Domain	64..106											
XX	Region	/label= Linker_region											
XX	Domain	107..471											
XX	Region	/label= Catalytic_core_domain											
XX	Domain	/note= "catalytic core domain is the preferred domain for use in constructs of the invention"											
XX	Region	WO9623928-A1.											
XX	Domain	08-AUG-1996.											
XX	Region	29-JAN-1996;											
XX	Domain	96WO-US00977.											
XX	Region	01-FEB-1995;											
XX	Domain	95US-0382452.											
XX	Region	(GENV) GENENCOR INT INC.											
XX	Domain	Clarkson KA, Collier KD, Fowler T, Larens E, Ward M;											
XX	Region	WPI; 1996-371466/37.											
XX	Domain	DR N-PSDB; AAT32221.											
XX	Region	Treatment of cellulose-contg. fabrics such as denim, e.g.											
XX	Domain	stone washing - using											

XX H: Insolens Cel16B fungal cellulase protein sequence.
 DE
 XX
 KM Cel16B: Cel16A: fungal cellulase; cleaning composition; conditioner;
 KM cellulolytically active protein; endo-1,4-beta-glucanase; enzyme;
 KM fabric softener; Humicola-like cellulase; glycosyl hydrolase family 6;
 KM detergent composition.
 XX
 OS Humicola insolens.
 XX
 PN MO9901544-A1.
 XX
 PD 14-JUN-1999.
 XX
 PF 02-JUL-1998; 98WO-DK00299.
 XX
 PR 04-JUL-1997; 97DK-0000813.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Andersen KV, Damgaard B, Lund H, Nielsen JB, Schulein M;
 XX
 DR WPI: 1999-106046/09.
 DR N-PSDB: AAX27947, AAX27948.
 XX
 PT Cleaning composition containing Humicola endo-beta-1,4-glucanase
 PT useful as detergent compositions or additives, or as fabric
 PT conditioners
 XX
 PS Claim 2; Page 247-249; 271pp: English.
 XX
 CC This sequence is the Humicola insolens Cel16B fungal cellulase.
 CC The invention relates to a cleaning composition (A) that contains at
 CC least one enzyme (I) with cellulolytic activity and at least 25 wt.% of
 CC total cellulolytically active protein present is Humicola
 CC endo-1,4-beta-glucanase or a Humicola-like cellulase of the glycosyl
 CC hydrolase family 6. (A) are used as detergent compositions or additives,
 CC or as fabric softener or conditioner. (I) provides colour clarification,
 CC and possibly soil removal, without significant weakening of cellulose
 CC materials during pre-soaking or wet storage (contrast use of enzymes from
 CC families other than 6).
 XX
 SQ Sequence 446 AA:

Query Match 20.0%; Score 449.5; DB 20; Length 446;
 Best Local Similarity 28.0%; Pred. No. 7.2e-30;
 Matches 120; Conservative 70; Mismatches 143; Indels 95; Gaps 16;

QY 6 NPVAGATFFVNPYMAOEVOG-----EAAQTATATLAKKRVV-STYSTAVWMDRIAIN 58
 DB 30 npsgrtllvnsdysakldtqrqflsrgdqn---aakkyqekvgtfymnsnflir 86
 QY 59 GVGNGGGLTLYLDALSSQOOGTTPVEVIEIYIDLPERDCALASNGELPRTAAGLOTYET 118
 DB 87 dld-----valqraaraakaargemp-lvglvlylprdcasagesseglklsqnlrnykn 140
 QY 119 QYIDPFIASISNPKYSLRIVTITIEPDSLPAVNTNMSIQACATAPVPEEGIEVALTKLH 178
 DB 141 eynvnpfagkik--aaedvqfavllepdalgnmvlgts-afcrnatrpqaeigylasqlq 197
 QY 179 AIPNVIYIMDAHSGWLPNNASGVYQVQKVLNASIGVNGIDGFEVTNTANTPLKEPF 238
 DB 198 a-shllhlyldvavgvlgvacklqptqevatlqkqgnaklirgfsnvsny----npy 252
 QY 239 MTATQOOGSPVESANRYQNNPDIENDYAVDLXSLVAAFGPSSIMLIDTLANGGGR 298
 DB 253 stmp-----ppytsg-----spspdeeryatnlanamrqrqlpqr----- 289
 QY 299 NEPTGPSTADVMTFVNQSKIDL---RQHRGLMCGNCGAGIGOPPOASPTDFPNAHDAY 355
 DB 290 -----lldqsrvalagarseswqvcnvpagfgqp---ftctnlnpnda1 332

QY 356 VWIKPEESDGTSAASDPTTGKKSDBPCDPTTYTTSYGLTNALPNSPIACQWFPAPFDOL 415
 DB 333 vvvkpgesdg-----gc-----gnggapaagmwfdagayqml 364
 QY 416 VAAARPAV 423
 DB 365 tqnahdei 372

RESULT 13
 AAM56739
 ID AAM56739 standard; Protein: 449 AA.
 XX
 AC AAM56739;
 XX
 DT 14-SEP-1998 (first entry)
 XX
 DE Orpinomyces cellulase celC.
 XX
 KM Cellulase; endoglucanase; cellobiohydrolase; celC.
 XX
 OS Orpinomyces sp. strain PC-2.
 XX
 PN WO9814597-A1.
 XX
 PD 09-APR-1998.
 XX
 PF 03-OCT-1997; 97WO-US18008.
 XX
 PR 04-OCT-1996; 96US-0027883.
 XX
 PA (UYGE-) UNIV GEORGIA RES FOUND INC.
 XX
 PI Chen H, Li X, Ljungdahl LG;
 XX
 DR WPI: 1998-240096/21.
 DR N-PSDB: AAV29472.
 XX
 PT New recombinant DNA encoding Orpinomyces cellulase protein - useful
 PT for, e.g. producing recombinant Orpinomyces cellulase in host cell
 XX
 PS Claim 1; Page 32-34; 69pp: English.
 XX
 CC This polypeptide comprises cellulase celC of Orpinomyces sp. strain
 CC OC-2, an anaerobic bovine rumen fungus Orpinomyces sp. PC-2. Its
 CC amino acid sequence was deduced from an isolated cDNA clone (see
 CC AAV29472). CelC has endoglucanase and cellobiohydrolase activity,
 CC with highest activity at pH 5.6-6.2 and 40 degC. CelA (see AAM56738)
 CC and celB (see AAM56742) cellulolytic enzymes of Orpinomyces sp. PC-2
 CC are also provided. Recombinant DNA molecules encoding Orpinomyces
 CC cellulase proteins are claimed, as well as recombinant cells
 CC selected from Saccharomyces cerevisiae, Escherichia coli,
 CC Aspergillus, Trichoderma reesei, Pichia, Penicillium, Streptomyces
 CC or Bacillus, and a method for producing recombinant cellulase by
 CC culturing these host cells.
 XX
 SQ Sequence 449 AA:

Query Match 19.5%; Score 439.5; DB 19; Length 449;
 Best Local Similarity 32.0%; Pred. No. 5.2e-29;
 Matches 136; Conservative 55; Mismatches 131; Indels 103; Gaps 18;

QY 2 THVDNPVAGATFFVNP-----YMAOEVOSEAAQNTATLAKKRVVSTYSTAVWMDRIAI 57
 DB 121 tlvnsdptsdnfflenelysnkktqgevdsqrgslqekakvkvpcaavlawsgat 180
 QY 58 NGVNGGGLTLYLDALSSQOOGTTPVEVIEIYIDLPERDCALASNGELPRTAAGLOTYE 117
 DB 181 nev-----arlynags-----kcvfvlylmiprdcnagsng-----gdadnltsg 223
 QY 118 TQYIDPFIASISNPKYSLRIVTITIEPDSLPAVNTNMSIQACATAPVPEEGIEVALTKL 177

D	b		224	-gynvnslyntlin--qpsnrvlmlleptdltgnlvtaaman-crivhmhbqalsyatsk	279
O	y		178	HALPNVYITDAASHGSLGWCPNNNAGSYVOEYOKVLINASIGVNGIDGFVTANTPTPIKEP	237
				: : : : : : : : :	
D	b		280	gtcknvrvyidaahggwl---nsadtraevlaeellrnaagkkrigdstvnshy-----	330
O	y		238	FMTATQGVGGQPRESANFYQMNPDIDEDVDYAVDLYSRLVVAAGFPSSIGMLIDLRLNGWG	297
D	b		331	-----qpvyse-yqyhqnlnra-----lestgv-----rgmkflvdtsrng---	365
O	y		298	PNEETGPSTATDVNTPEVNOSKIDLRQHNGLMCNONGAGIAGOPPOASPDPDPMNH-LDAYV	356
				: :	
D	b		366	-----rnpssat-----wcnlkgaqlgarprapnd-pmnrplday	400
O	y		357	WIKPRGSDDTSAASDPPTTGKKSDPCMDPTTYTSGLVTNLALNSPIAGOWFPAQEDQLV	416
				: : : : : : : :	
D	b		401	wikpgdsdsassa-----dvcrn-----sdslgagaapaagwfndyfwmll	442
O	y		417	ANAP 421	
D	b		443	enamp 447	
 RESULT 14 AAB18414 ID AAB18414 standard; Protein; 432 AA.					
A	C		AAB18414;		
X	X				
D	T		15-JAN-2001 (first entry)		
X	X				
D	E		Amino acid sequence of a Celf cellulase protein.		
K	M		Celf; cellulase; endoglucanase; cellobiose; glucose; cellobiose;		
K	M		cellulosic material; textile; paper; ethanol; fuel;		
K	M		cellulosic feed stock.		
X	S		Orpinomyces sp.		
X	X				
F	H		Key	Location/Qualifiers	
F	T		Peptide	1..21	
F	T			/note= "signal peptide"	
F	T		Protein	22..432	
F	T			/note= "mature protein"	
F	T		Domain	22..57	
F	T			/note= "fungal cellulose binding domain"	
F	T		Peptide	67..105	
F	T			/note= "Asn-rich linker sequence"	
F	T		Domain	106..432	
F	T			/note= "catalytic domain"	
X	N		US6114158-A.		
X	N				
P	D		05-SEP-2000.		
X	X				
P	F		17-JUL-1998:	98US-0118319.	
X	X				
P	R		17-JUL-1998:	98US-0118319.	
X	X				
P	A		(UYGE-) UNIV GEORGIA RES FOUND INC.		
X	X				
P	I		Lt X, Ljungdaht LG, Chen H;		
X	X				
D	R		WPI; 2000-593548/56.		
D	R		N-PSDB; AAA75403.		
P	T				
X	X				
P	S		Disclosure; Column 13-16; 26pp; English.		
X	C		The present sequence represents a Celf protein of Orpinomyces sp. PC-2		

[illegible]

PI Chen H, Li X, Ljungdahl LG;
 XX
 DR MPI; 1998-240096/21.
 DR N-PSDB; AAV29471.
 XX
 PT New recombinant DNA encoding Orpinomyces cellulase protein - useful
 for, e.g. producing recombinant Orpinomyces cellulase in host cell
 XX
 PS Claim 1; Page 28-30; 69pp; English.
 CC This polypeptide comprises cellulase cels of Orpinomyces sp. strain
 CC OC-2, an anaerobic bovine rumen fungus Orpinomyces sp. PC-2. Its
 CC amino acid sequence was deduced from an isolated cDNA clone (see
 CC AAV29471). Cella has endoglucanase and cellobiohydrolase activity,
 CC with highest activity at pH 4.8 and 50 degC. Cells (see AAV56742)
 CC and cels (see AAV56739) cellulolytic enzymes of Orpinomyces sp. PC-2
 CC are also provided. Recombinant DNA molecules encoding Orpinomyces
 CC cellulase proteins are claimed, as well as recombinant cells
 CC selected from Saccharomyces cerevisiae, Escherichia coli,
 CC Aspergillus, Trichoderma reesei, Pichia, penicillium, Streptomyces
 CC or Bacillus, and a method for producing recombinant cellulase by
 CC culturing these host cells.
 XX
 SQ Sequence 459 AA;

Query Match 18.3%; Score 411; DB 19; Length 459;
 Best Local Similarity 30.0%; Pred. No. 1.5e-26;
 Matches 124; Conservative 66; Mismatches 130; Indels 94; Gaps 20;

QY 13 FEVNPYMAQ-EVQSE---AANQTNATLAAMRVSTYTAVMMDRIAINGNGP-GLT 67
 DB 133 ffeneIysnykfgevdIsIkkingdIkakekvkyptavwl-----awdgapevp 185
 QY 68 TYDPAALSOOGQTPREIEIYVDLPBGCAALASNGELPATAGLOTYETQYIDPIASI 127
 DB 186 ryl-----geagn--kcvfvlymIprcdganasag---gsatIdkyk-gyInnIyn- 232
 QY 128 LSNPKYSSLRIYVIEIPDSLPNAVNTNMSIQACATAVPEYEGIEYALTKLAIPNYIYM 187
 DB 233 -tanqykneklvmllepdtlglnvtinn-dncnvrnmhbkqalskfgtshvkvyl 290
 QY 188 DAAHSGMLGMPNNAAGVQEVQKVLNASTGVNGIDGFTVNTFANTYPLKEPPMTATQOVGG 247
 DB 291 daahgawln--gyadqtaavIkelln-naagskIrgIsItnvsny----- 331
 QY 248 QPVESANFYQMNPDIIDEADYAVDLSRLVAAGFPSSIGMLIDILRNCGMGPNPTGPSTA 307
 DB 332 qstese--ykYhqnlna-----leskyv-----rgIkIIdtsrng----- 366
 QY 308 TDVNTFEVNOGKIDLRQHGLMCNONGAGLGQPOASPTDFPNAHLDAYVMIKPPGESDGT 367
 DB 367 anvegafnas-----gtwcnfksgaglgqrpkgnpnsmpIldaymwIkcpgeadgs 418
 QY 368 SAASDPTGKKSQPCDPYTTTSYGVLTNALPNSPIAGQWFPAPQDQLVANARP 421
 DB 419 s-----qgaradpvc-----argdsIlgapadagswfheyftmlIqnanp 457

Search completed: August 29, 2002, 16:21:42
 Job time: 317 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:24:01 ; Search time 72.78 Seconds

(without alignments)
558,475 Million cell updates/sec

Title: US-09-917-384-4

Perfect score: 2249

Sequence: 1 ATHVDNRYAGATFEVNPYWA.....AGQWPAQFDQLVANARPAV 423

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	1244	55.3	872	2	S49541
2	1171	52.1	683	2	A82704
3	1050	46.7	579	2	T35240
4	999	44.4	596	2	A55976
5	639.5	28.4	438	1	S70602
6	546	24.3	471	1	A26160
7	546	24.3	471	1	A38979
8	330.5	14.7	380	2	G70847
9	261.5	11.6	441	2	T12011
10	244.5	10.9	449	2	A24993
11	242.5	10.8	426	2	A42360
12	193.5	8.6	359	2	JN0544
13	143	6.4	388	2	JT0308
14	120	5.3	1180	2	E86719
15	113.5	5.0	693	2	D90441
16	113	5.0	638	2	G97171
17	112	5.0	329	2	G82900
18	111.5	5.0	1131	2	T41144
19	110	4.9	2348	2	AD1841
20	108.5	4.8	745	2	G72453
21	107	4.8	657	1	A64079
22	107	4.8	728	1	A38084
23	105.5	4.7	1217	1	G18177
24	103.5	4.6	517	1	G18177
25	103	4.6	583	2	G90327
26	103	4.6	1268	2	B99789
27	103	4.6	1270	2	E85649
28	103	4.6	3029	2	S76109
29	103	4.6	3972	2	S75251

ALIGNMENTS

Query Match	Score	DB 2	Length	872
Best Local Similarity	55.78	Pred. No. 2.1e-63		
Matches 239	Conservative	60	Mismatches 120	Indels 10
			Gaps	6
3	HVDNRYAGATFEVNPYWAQVSEANQNT-NATLAKMRVSTFVMMDRIAINGVN 61			
44	HVDNRYAGATFEVNPYWAQVSEANQNT-NATLAKMRVSTFVMMDRIAINGVN 103			
62	GGEPLTLYTDAALSOQGT-TPEVIEIYVDLPGRDCAALASNGELPATAGLQTYETQY 120			
104	DGNGLKRLHNDNAVQKACAGVPLVFNLYVDLPGRDFALASNGELPATAGLQTYETQY 163			
121	IDPFIASLSPKYSLSLIVITTEPDSLPNAVNTNNSIOACATVYVYEGEYATLTKLHAI 180			
164	IDPFIADLDDPEYESTIRIAITTEPDSLPNAVNTNNSIOACATVYVYEGEYATLTKLHAI 223			
181	PNYIYVYDAASGLWGLPNNNA---SGYVOEYOKVLNASIGVNGIDGFTVTANTPTPEK 237			
224	PNYIYVYDAASGLWGLPNNNA---SGYVOEYOKVLNASIGVNGIDGFTVTANTPTPEK 281			
238	FMT-ATQVGQAPVESANFYQWNPDIADYAVDLYSLVAAGFPSSIGMLIDTLRNGWG 296			

probable bacteriophage
hypothetical prote
hypothetical prote
zinc metalloprote
F420-nonreducing h
cellulase CelE or
probable ppe prote
N-acetylmuramyl-L
cellulase (EC 3.2.
proteoglycan core
beta-1, 3-glucanas
SHC1 protein - yea
parasporal crystal
probable membrane
differentiated ker
probable exported

Db 282 ILSDSLTINTNTPIRSSKFTYMNFDDEIDYTAHHHRLLYAAGFPSSIGMLVDTSRNGW 341

Qy 297 GNEPPGPSTATDVNTFYVNSKIDLRHGRGLMCNONGAGLGOPPOASPFDPNHAHDVY 356

Db 342 GPNRRPSTASTSDVNAVDNARVRRVHRCAMCNPLGAGIGRFPPEATPGPASHLDAFY 401

Qy 357 WIKRPESDGTSAASDPPTTGKSDPMCDPTYTTS--YGVLTNALPNSPIAGCMFPNPFQ 414

Db 402 WIKRPESDGTSAASDIPNDGGRKFRMCDPTFVSPRLNOLTGATPNAPLACGMFEEQFVT 461

Qy 415 LVNARPAV 423

Db 462 LVNARPAV 470

RESULT 2

A82704

1.4.beta-cellobiosidase XF1267 [imported] - Xylella fastidiosa (strain 9a5c)

C.Species: Xylella fastidiosa

C.Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C.Accession: A82704

R.anonymus, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequence

Nature 406, 151-157, 2000

A.Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A.Reference number: A82315; MUID:20365717

A.Note: for a complete list of authors see reference number A59328 below

A.Accession: A82704

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-683 <SIM>

A.Cross-references: GB:AE003960; GB:AE003849; NID:g9106242; PIDN:AAF84076.1; GSPDB:GN001

R.Experimental source: Strain 9a5c

R.Simpson, A.J.G.; Relnach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Bionesi, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre, R

as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincanli, A.P.; Ferreira, A.J.S.

submitted to Genbank, June 2000

A.Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Jungueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, B

A.Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri, D.A.

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa

A.Authors: de Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silve

M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; V

A.Reference number: A59328

A.Contents: annotation

C.Genetics:

A.Gene: XF1267

Query Match 52.1%; Score 1171; DB 2; Length 683;

Best local Similarity 53.0%; Pred. No. 3.4e-78;

Matches 227; Conservative 59; Mismatches 130; Indels 12; Gaps 6;

Qy 3 HUNNPYGAFFVVPYNAOEVOSEANOTNATLAAMKRVSTYSTVWMDRIAINGVNG 62

Db 21 HUNNPEFADASTYLPDYSKIDTSTIAKANDVSLKERMQIRKYPPTGWMVLDITDAVVGPK 80

Qy 63 GPC--LTFYIDALALSOOGQTPEVIEIVYDLPKDCALASNGELPATPAAGIQYETO 119

Db 81 NPGRLNLTGLHDAALAKKGMPTTATFYIYIDIPGRDCAALASNGELPLTPEGLQKRE 140

Qy 120 YIDPIASILSNPKYSLRIVTIIIEPDSLPMNAVTNNSIOACATA--VPYEOGLEVALTKL 177

Db 141 YIDTIASIFSMPKXKDIRIVNIIIEPDSLPIITNNSTPSCAAKKNGIIEEGIKYALNKL 200

Qy 178 HAIPNVIYMDAHSGLGWNNASGYVOVKL--MASIYNGIDIGFTYNTANTPYLKE 236

Db 201 SEIPNVIYMDIGHSGLGWNTNRIPAVSLYTKYIOSTIAGASVNGFNTANTPYLIE 260

Qy 237 PFM-TAQOOGGQGVESANFYOWNPDDIDADYAVDLSYLVAAGPSSTIGMLDITLNGW 295

Db 261 PNLPNPLNTGGQVIRSSKFTYMNRYFDEMDSYETLYNDPVAAGWSSSGITIIDTGRNGW 320

RESULT 3
 T35240
 Probable secreted cellulase - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
 C:Accession: T35240
 R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 Submitted to the EMBL Data Library, September 1998
 A:Reference number: Z21572
 A:Accession: T35240
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-579 <SEP>
 A:Cross-references: EMBL:AL031515; PIDN:CAA20645.1; GSPDB:GN00070; SCOEDB:SC5C7.33
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SC5C7.33

Query Match 46.7%; Score 1050; DB 2; Length 579;
 Best Local Similarity 49.8%; Pred. No. 2,1e-69;
 Matches 214; Conservative 61; Mismatches 125; Indels 30; Gaps 8;

QY	4	VDNFPAAGTFFVNPYMAOEVSAAANTATLAAMRVSTYSTAVWMDRIATGNGG	63
DB	165	VDNFPAAGQVYVNPYPMASNAAPGDR-----IADEPTGVWLDRIATGANGG	214
QY	64	PGLTYYDAAALSOQGGTTPVEIYIYIDLPGRDCALASNGELPATAGLQTYETQYIDP	123
DB	215	MLGRDLDAAL-EKSGSEMYVQVLYIYMLPGRDCSALASNGELGPT--IDRYKTEYIDP	271
QY	124	IASLISNFKYSIRIVTILIEPDSLPNATNNISQACAT-----AAYVYEQGLEIATLK	176
DB	272	IAEILISDSKIVADLRIVTVTVEIDSLPNLTANNVSGRPTAENCNDYMKANGNYQKGAYALNR	331
QY	177	LHAIPNVYIYMDAAHSGWLGMPNNASGYVOYQKYLNASIG--VNGIDGFYTNANTYPL	234
DB	332	LGAVGNYVNYVDAGHGHGLGWDN--FGSALEIFKRAATBEATGLDDVNGEFLVNTANYAL	390
QY	235	KEPMTATQOYGGORVESANFYQWNPDIIDEADYAVDLTSLVAAGFPSSIGMLIDTLRNG	294
DB	391	KEENFKIDSNGTSTSVRSQDWVWQYDELSYQAMDKYLSLGFQDNLMLIDTSHNG	450
QY	295	WGGPEPPTGPATPVNFVFNOSKIDLQHHGLMGNONGAGIGOPPOASPDPEFAHIDA	354
DB	451	WGGADRPFGPGATTDVNTYVNGGRIDRIHILGNMCCNOSGAGIGEPQASPA---AGIDA	506
QY	355	YVWIKPGEISDGTSAASDPTTKKSDPKCDPTYT--TSYGITNALNSPIAGQWFPQDL	411
DB	507	YVWMMKPGESDSGSKLIDNPDGKGFDRMCCDPTTYIGENNGMSGALDPAISGAMFSAQ	566
QY	412	FDQALVANRP 421	
DB	567	FOELMKKNAYP 576	

RESULT 4
 T353976
 Cellulose 1,4-beta-celllobiosidase (EC 3.2.1.91) - Thermomonospora fusca

C:Species: Thermomonospora fusca
C:Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 22-Oct-1999
C:Accession: A55976
R:Zhang, S.; Lao, G.; Wilson, D.B.
Biochemistry 34, 3386-3395, 1995
A:Title: Characterization of a Thermomonospora fusca exocellulase.
A:Reference number: A55976; WUID:95186496
A:Accession: A55976
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-596 <ZHA>
C:Cross-references: GB:018978; NID:9664822; PIDN:AAA62311.1; PID:9664823
C:Superfamily: bacterial cellulose-binding domain homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
E:40-140/Domain: bacterial cellulose-binding domain homology <BCB>
F:41-139/Disulfide Bonds: #status predicted

Query Match	44.48;	Score 999;	DB 2;	Length 596;
Best Local Similarity	48.38;	Pred. NO. 1.2e-65;		
Matches 209; Conservative	55;	Mismatches 137;	Indels 32;	Gaps 9

```

Oy      4 VDNPAAGTFFVNPWMAEVOESAAQNQNTLAAMKRVYSTRVAVWMDRTAATNG----- 59
Db      179 VDNPEBGGKLLVNPWMSKAAAEPEGG-----AVANESTRVAVWLDRTGALTEGNDSP 228
Oy      60 VNGGPGTLTYVLDALASQOOGTTPVEIEIYVYDLDEGRDCALASNGELPATAAGLOTYETQ 119
Db      229 TTGSGAGLNDHLEEVN-RQSGDPLTIGVYIYNLPRDCCALASNGELGPDE--LDRYKSE 285
Oy      120 YIDPASTLISLN-PKYSLSRIYTIIEPDSLPAVAVNMTSIQACATAVPY-----YEOGIEY 172
Db      286 YIDPDIADIMMPEFADENRIVAIIEIDSLPMLVYNVWGSGGTETCALAYMKONGGVYNGVGY 345
Oy      173 ALTKHAIHPNYIYVYDAHSGMLGMPNNAAGYDEOVOKVLNA--IGVNGIDEPYNTANY 231
Db      346 ALRKIGELIPNYIYNITDAAHHGHIHGDMSNFGSVDIIFYAANAAGSTVDYVIGCFISNTANY 405
Oy      232 TPLKEPENTATQOVGGQPVESANFYQWNPDPIDEADYADVLSRLVAAGPSPSIGMLIDTL 291
Db      406 SATVEPYLDVNGTVNGQLIRQSKWMDWNYUDELSEFYDDLQALILAKFRSDIGMLIDTS 465
Oy      292 RENGSGEPBEPNGSPATVNTVFVNSKIDLQHRGLMCONGAGLGGORPASPIPDFEPNAN 351
Db      466 RENGSGEPBEPNGSPSSSTLNTLNTYVDSRIDRKRIHPEMNCNQAGAGIGERPYVNPV---PG 521
Oy      352 LDAYVWIKPESBDGTSAAASPTTGKKSKSDPPMCDPTTYTTSYVLTN---ALNPSSIAGOMF 408
Db      522 VDAYVWVAKPESBDGASASEIPENDECKGFRMCDPTYGQANANGNPNNSCALNAPISGIMF 581
Oy      409 PAQFDQVLAANRP 421
Db      582 SAQFRELLANRP 594

```

RESULT 5
S70602
cellulose 1,4-beta-cellulohydrolase (EC 3.2.1.91) II precursor - cultivated mushroom
N:Alternate names: cellulase
C:Species: Agaricus bisporus (cultivated mushroom)
C:Date: 10-Sep-1999 #sequence,revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: S70602
R:Yaguee, E.; Chow, C.M.; Challen, M.P.; Thurston, C.F.
Curr. Genet. 30, 56-61, 1996
A>Title: Correlation of exons with functional domains and folding regions in a cellulase
A:Reference number: S70602; MUID:9626930
A:Accession: S70602
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-438 <YAC>
A:Cross-references: EMBL:Z34007; NID:g1494969; PIDN:CAA83971.1; PID:g1494970
C:Genetics:
A:introns: 34/1; 56/2; 96/3; 202/1; 325/3; 410/2

C:Superfamily: cellulose 1,4-beta-cellobiosidase II; fungal cellulose-binding domain
 K:Keywords: glycosidase, hydrolase, polysaccharide degradation
 E:25-36/Domain: fungal cellulose-binding domain homology (FCB)

Query Match	28.48;	Score 639.5;	DB 1;	Length 438;
Best Local Similarity	36.68;	Pred. No. 2.2e-39;		
Matches 153;	Conservative 62;	Mismatches 134;	Indels 69;	Gaps 12;

```

0Y 6 NYAAATEFYVNAOEYOESEAQNTNATLAAKMRVSTYSYSAWMDIAIANGVSGPG 65
    ||| | : : : : ||| | : : ||| | : : ||| | : : ||| | : :
Db 90 NYTKRTWLSFYZADEVAQAADISNPSLTKAASAKITPEYWFDDYAVY-----PD 143
    ||| | : : : : ||| | : : ||| | : : ||| | : : ||| | : :
0Y 66 LTTYDAALSOQOQCTTPEVIEIYIDLPGRCCALASNGELPATAGALOTETGYDPIA 125
    ||| | : : : : ||| | : : ||| | : : ||| | : : ||| | : :
Db 144 LGGYIADARSKNQ-----LYQIVYVDLPDRCCALASNGEESLNDGJMKTK-NTYQDIA 197
    ||| | : : : : ||| | : : ||| | : : ||| | : : ||| | : :
0Y 126 SILSNPKYSSLRIYITIEPDSLPNAVYTMKMSIOACATAPYEEOGIEYALTYLHAIPNYI 185
    : : : : : : ||| | | : : : : ||| | : : ||| | : : ||| | : :
Db 198 AQIK--QEPDVSVAIVIEPDSLANLYTMJNVOCKCANASAKKEGYIVAVOKLNAV-GVTM 254
    : : : : : : ||| | | : : : : ||| | : : ||| | : : ||| | : :
0Y 186 YVDAHSHGMLGPNNAASGYOEVOKVLNASIGVNGIDGFNTYATYTLKKEPMATQOY 245
    ||| | : : ||| | | : : | : : : : ||| | : : ||| | : : ||| | : :
Db 255 YIDACHAGLGPANLSPAAQLFAQIYRDAGSPNNLRKIATVAFNALRAS----- 306
    ||| | : : ||| | | : : | : : : : ||| | : : ||| | : : ||| | : :
0Y 246 GQOPESANFYQWNPNDIDEADYADVLSRLVAAGFPSSIGMLIDLRLMGWCGPNEPTGPS 305
    | : : : : : ||| | | : : : : ||| | : : ||| | : : ||| | : :
Db 307 SPDPIT-----QGNSEYDEIHYTEALPMLSNAGPRAN--FYDQGRSG----- 348
    ||| | : : ||| | | : : | : : : : ||| | : : ||| | : : ||| | : :
0Y 306 TATDVNTFVNSOKIDLKQHRGLMCNONGAGLQOPQASPTDFPNAHLDAYVWIKRPGESD 365
    : : : : : : ||| | | : : | : : : : ||| | : : ||| | : : ||| | : :
Db 349 -----VONIRIQMGDMCWKAGGAGQGRP---TTWNGSSLIDAIWVVRPGCECD 393
    ||| | : : ||| | | : : | : : : : ||| | : : ||| | : : ||| | : :
0Y 366 GTSASADPTTGKKSPKMDPTTYSYGVLYTNALPSPITAGQWFRPQOLYANARPAY 423
    ||| | : : ||| | | : : | : : : : ||| | : : ||| | : : ||| | : :
Db 394 GTSNDSSP---RFSHCS-----LSQDHQPARPAGTGWFAAYFETLYAANNPAL 438
    ||| | : : ||| | | : : | : : : : ||| | : : ||| | : : ||| | : :

```

RESULT 6
A26160
cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) II precursor - fungus (*Trichoderma reesei*)
N:Alternate names: exo-cellulobiohydrolase II
C:Species: *Trichoderma reesei*
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A26160; A26472
R:Chen, C.M.; Gritzall, M.; Stafford, D.W.
Bio/Technology 5, 274-278, 1987
A:Title: Nucleotide sequence and deduced primary structure of cellulobiohydrolase II from *Trichoderma reesei*
A:Accession: A26160
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-471 <CHE>
R:Teeri, T.T.; Lehtovaara, P.; Kauppinen, S.; Salovuori, I.; Knowles, J.
Gene 51, 43-52, 1987
A:Title: Homologous domains in *Trichoderma reesei* cellulolytic enzymes: gene sequence
A:Reference number: A26472; MUID:87248061
A:Accession: A26472
A:Molecule type: DNA
A:Residues: 1-471 <TEB>
A:Cross-references: GB:M16190; NID:g170540; PIDN:AAA34210.1; PID:g170541
C:Genetics:
A:Introns: 31/2; 160/2; 243/1
C:Superfamily: cellulose 1,4-beta-cellulobiosidase II; fungal cellulose-binding domain
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:31-62/Domain: fungal cellulose-binding domain homology <FCB>

```

Query Match      24.3%; Score 546; DB 1; Length 471;
Best Local Similarity 33.6%; Pred. No. 1.9e-32;
Matches 145; Conservative 51; Mismatches 154; Indels 82; Gaps 13;

OY 1 ATHVNPPIAGATFEVNPPIWAOEVSSEANQNTATLAARKRVSTVSTAVMD----- 52

```

[illegible]

RESULT 7
A38979
cellulose 1,4-beta-celllobiosidase (EC 3.2.1.91) II - fungus (*Trichoderma viride*)
C:Species: *Trichoderma viride*
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A38979
R:Wang, J.R.; Zhang, M.F.; Huang, T.
Acta Genet. Sin. 22, 74-80, 1995
A>Title: The primary structure of cellobiohydrazase gene (CBH II) from *Trichoderma viride*
A:Reference: A38979
A:Accession: A38979
A:Molecule type: DNA
A:Residues: 1-471 <MAN>
C:Genetics:
A:Gene: cbhII
A:Introns: 31/2; 160/2; 243/1
C:Superfamily: cellulose 1,4-beta-celllobiosidase II; fungal cellulose-binding domain homolog
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
C:31-62/Domain: fungal cellulose-binding domain homology <FCB>

Query Match	24.3%;	Score 546;	DB 1;	Length 471;
Best Local Similarly	33.6%;	Pred. No. 1.9e-32;		
Matches 145;	Conservative 51;	Mismatches 154;	Indels 82;	Gaps 13;

[illegible]

```

QY      231 YTPLEKPEPMATATQOQGGVPSSANFYQWNPIDIDADYADVILSLVLAAGPSSIGMLIDT 250
Db      330 YN-----GWNITSPPTQGNNAVYNEKLYIHAGPLLANHGW-SNAFFTTDQ 375/5
QY      291 LKNGMGGENEPTGPSTATDVNTFVQSKIDLRQHRGLMCNONGAGLQGPQASPTDFPNA 350
Db      376 GMSG-----KQPLTG-----QQQMGDMCNVIGIGFIRSAMTGD---S 410
QY      351 HLDAYVWIKIPROESDGTSAADPTTGKSKSDPMCDPTTYSIGVLTNALPNSPIAGOWEFA 410
Db      411 ILDSFVWVWKPGEGECGTSDSSAP-----RFDShc-----ALPDALQAPQAPAGAMFQA 457
QY      411 QFDOLVANRA 422
Db      458 YFVQLLTANAPS 469

```

RESULT 8
 G70847
 probable cellulase - Mycobacterium tuberculosis (strain H37Rv)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: G70847
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
 ; Connor, R.; Davies, R.; DeLisi, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd,
 R.; Janczewska, J.; Jones, S.; Kaul, R.; Kiehl, J.; Martin, J.; McAnulty, J.; Mouton,
 R.; Ndindam, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellom, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Accession number: A70500; MUID:98295987
 A:Reference: G70847
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-380 <COL>
 A:Cross-references: GB:AL021428; GB:AL123456; NID:g3261514; PIDN:CAA16243.1; PID:e12424
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: celA

Query Match	14.78	Score 330.5	DB 2	Length 380
Best Local Similarity	29.18	Pred. NO. 9.7e-17		
Matches 109	Conservative 42	Mismatches 123	Indels 101	Gaps 18

```

QY      6  NPNACATFEVNPYMAQEVQSEMAQNTANTLAAKRRVSTYSTAVMDRIALINGVNGPG 65
Db      89  NPLACKRPYYVDPASAAVMAARNAAPNAELT SVANTPOST ---WLDQ -AFPPATVGT 142
QY      66  LTTYDALDSOOGCTTPEVIEIVYIDLPGRPCALASNGELPATPAALQIYVEYQYIDPIA 125
Db      143  VARYTGAA -QAAGAMP---VLTIXGPHRCCGYSAGG---PATG -TDYKWDIDAVA 191
QY      126  STL-SNPYSSLRITVITIEPDSLPRNAVNTMSIOACATAVPYEEOG--LEYALTKLHAIPN 182
Db      192  SGLGSSP-----ATIIYEPDAL-----AMADCLSPDQOEFEFLVYRAVDTLTRDPA 238
QY      163  YTIYDAAHSGHLGHPNNAAGSVQEVQYKVLNASTIGVNGIDGFVTYNTANTYPLKEPEKTAT 242
Db      239  AAVYVVDACHSRLS-----AEMAMARLN-DVGGVGRARGFSLNVSN-----FYTTD 282
QY      243  OOVGQAPVESANFYQWNPDIIDEADYAVDLYSRLLVAGGPPSIGMLITLRLNGWGNPREPT 302
Db      283  EETIG-----YEAL-SGTLNASHYIDITSRNG-----A 309
QY      303  GPSTATDVNTFVNSKIDLRQHRGLMCQONAGLIGOPPOASPTDFPNAAHDAVYMIKPG 362
Db      310  GPAPDAPLN-----WCNPSRRAIGARP---TTTAAAHADATVIMIKRPG 350
QY      363  ESDGTSAAADPTTGK 377
Db      351  ESDGTGCGGEPQAGR 365

```


Matches	106;	Conservative	44;	Mismatches	104;	Indels	129;	Gaps	24;
QY	5	DNPYAGATFFVNP-----YMAQVQSEANQTNATLAAKMRVSYSTPAWMDRIAAT-N	58						
Db	33	DSP-----FYVNPNNSSAEWVR-----NNNDP-----RPVINDRIASVQ	69						
QY	59	G---VNGSPG-LTTYLDAAALSQGGTTEVEIVETIYDLPGRCCALASNGELPATPAAGQ	114						
Db	70	GTWFANHNPQGTGGVDALMSAAQ-AAGKIPILVYMAGRGCRGHHSSGQ-----APSHS	123						
QY	115	TYETQYIDPIASILSN-PKYSRLRIVTITIEPDSLPAWVYNNMSIQACATVPRYEQI---	170						
Db	124	AYRS-WIDEPAAAGLKNRRAY-----IIVERP-----LISMSSCMOHVQDEVLET	167						
QY	171	EYALFKLHA-IPNVYITMDAAHSCWLGMPNNASGEVQEVORVLNASIGVNGIDGFVTNT	228						
Db	168	MAYAKAKAKAGSSQARIYTFDAGHS-----ASDSPQQMASMLQOADISNSAHGATGAT	219						
QY	229	ANYTFLEKPEPMHATQOVVGQPYESANFYQWNNDDIDEADVLYDLSRLVAAGPSSIGMLI	288						
Db	220	SN-----TRMT--ADEVAALAKAVLS--AIGNP-SLRAYI	248						
QY	289	DTLRNGMGSPNEPTGPSTATVDYNTFVNQSIDLRQHRGILCMQNGANAGLGPQOASPSTDP	348						
Db	249	DTSRNG-----NSPAG-----NE-----WCPSPSGRAIGTP---STTNG	279						
QY	349	NAHLDAIYVIRKPGESDGTSAAS	371						
Db	280	DPMTDAFLMIKIRPGEADGCIAGA	302						

```

RESULT      12
JN0544      cellulase (EC 3.2.1.4) casa - Streptomyces sp.
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Streptomyces sp.
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Oct-1999
C:Accession: JN0544; S19688
R:Dammund, H.G.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.J.
Gene 123, 105-107, 1993
A:Title: Endoglucanase Casa from alkalophilic Streptomyces strain KSM-9 is a typical member of the family 1A
A:Reference number: JN0544; MUID:9318417
A:Accession: JN0544
A:Molecule type: DNA
A:Residues: 1-359 <DAM>
A:Cross-references: GB:I03218
A:Experimental source: Strain KSM-9
R:Gilkes, N.R.; Claessens, M.; Neherold, R.; Hentisat, B.; Meinke, A.; Morrison, H.D.
Eur. J. Biochem. 202, 367-377, 1991
A:Title: Structural and functional relationships in two families of beta-1,4-glycanases
A:Reference number: S19688; MUID:92104156
A:Accession: S19688
A:Molecule type: DNA
A:Residues: 185-222 <GTD>
A:Cross-references: EMBL:X61008; NID:947520; PIDN:CAA4330.1; PID:947521
C:Genetics:
A:Gene: casa
A:Start codon: GTG
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel-
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match      8.6%, Score 193.5; DB 2; Length 359;
Best Local Similarity 24.7%, Pred. No. 1.1e-06;
Matches 82; Conservative 38; Mismatches 107; Indels 105; Gaps 18;

OY      42 VSTYTAVMORIRAIINGVNGSGLTLYLDALISQGGQTTPVEIVIVYIDLCRCACALA 101
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      107 IGSQDAWF--AGAYN-----PGTITQVAEVISAAAAAGGLPVVPMIDPFRCCGN-H 158
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      102 SNGELPATNA-----GLQTYETQYIDPTASILSNPKYSLSRLVITIIEDSLP--NAVTNM 154

```

```

Db      155  SGGAGSPFAAAYEMSGLA-----AGLGSE-----VVVLEPDAIPLTICDINO 203
Qy      155  SIQACATAVPYEEQIEYALTKHAIPNVYLYMDAAHSGMLGMPDNASGYVOEYOKVLA 214
Db      204  QQAERIALD-----AGLAEAVT--DANPEARYYYVGHSAW-----HAPAAI--APTLYEA 250
Qy      215  SIGVNGIGEFVNTNATPTLKEPEFATAOQGGOGPVEASNFYQWNPIDEDAVDYKSR 274
Db      251  GILEHGA-GIANISINY-----RTTIDELTA-----YAS 277
Qy      275  IYAAQFPSSIGLIDTLTNGMCGNPEPSPSTALDVNTFVQSKITDLRHRGLMCNONGA 334
Db      278  AVIAELGGGLGAVVDTSRNG-----NGFLGSE-----WCDDPGR 311
Qy      335  GIGQPPQASPTDFNAHLDAYVIKKPGESEDG 366
Db      312  LVGNMPTVP--GVPG--VDAAFLWIKLPGELDG 340

```

```

RESULT      13
JT0308
cellulase [EC 3.2.1.4] I precursor - Streptomyces sp. (strain KSM-9)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Streptomyces sp.
C.Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 22-Oct-1999
C.Accession: JT0308
R:Nakai, R.; Horinouchi, S.; Beppu, T.
Gene 65, 229-238, 1988
A>Title: Cloning and nucleotide sequence of a cellulase gene, casa, from an alkalophilic bacterium, Nitrospira helveticus strain 172
A.Reference number: JT0308; MUID:88313671
A.Accession: JT0308
A.Molecule type: DNA
A.Residues: 1-388 <NA>
A.Cross-references: GS:M20921; NID:g153340; PID:AAA26776.1; PID:g153341
A.Note: part of this sequence, including the amino end of the mature protein, was compared with the sequence of the mature protein from Nitrospira helveticus strain 172.
C:Genetics:
A:Gene: casa.
A:Start codon: GTG
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as amylose, glycogen, dextran, and cellulose.
A:pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1.70/Domains: signal sequence #status predicted <SIG>
F:71-388/Product: cellulase #status experimental <MAT>

Query Match          6.4%; Score 143; DB 2; Length 388;
Best Local Similarity 22.2%; Pred.No. 0.0063;
Matches 86; Conservative 40; Mismatches 111; Indels 150; Gaps 20;

QY 48 AYWMDRIAINGVNGSGPLITLTDALASOOCTFEVEIEIYI-YDLPGDCAALASNEL 106
    ||| | | | | | | | : : : : : : : : : : : : : : : :
DB 113 AWF--AGAYN-----PGLTIOVAEIVTSRRQPQGPVYPYPIPRDCGN-HSGGA 164
    ||| | | | | | | | : : : : : : : : : : : : : : : :
QY 107 PA-----TAAGLQTFEQYIDPIASILSNPKYSLSLRVTIIEDSLPN--AVTN 153
    | : | | | | | | | : : : : : : : : : : : : : : : :
DB 165 PSFAAYAEWSGLEAFACLG--EPVVVYLSPMRFWI-----DCLENQRARERL 210
    ||| | | | | | | | : : : : : : : : : : : : : : : :
QY 154 MSIQACATNAVPEYEGEIEVALTKLHAIPNVYIYMDAAHSGLGWENNASGYVQEVKYN 213
    :::: | | | | | | | | : : : : : : : : : : : : : : : :
DB 211 AALQASPEAV-----TDANPEARLYYDVGHSAW----HAPPAI--APTIVE 250
    ||| | | | | | | | : : : : : : : : : : : : : : : :
QY 214 ASIGVNGIDGFPTNTANTYPLKEPEMTAQOQGQPVESANFYQNPNPIDEDADVADLYS 273
    | | | | | | | | | | : : : : : : : : : : : : : : : :
DB 251 AGILEHGA-GIAFTINSNY-----RTTTDETA-----YA 277
    | | | | | | | | | | : : : : : : : : : : : : : : : :
QY 274 RLVAAGFPSSIMLIDTLERNWGNGNEPTGSPSTATDVNTFVNQSKIDLROHRLMCNQNG 333
    | | | | | | | | | | : : : : : : : : : : : : : : : :
DB 278 SAVIAELGGGLCAVADITSNG-----NGPIAALVNI----- 309
    | | | | | | | | | | : : : : : : : : : : : : : : : :
QY 334 AGIGPPQASPTDFNFAMLDAYVIWKIPGEISDGTSAASDPPTTGKKSDMCDPYTTSYG 393
    | | | | | | | | | | : : : : : : : : : : : : : : : :

```

Db 310 -----RTVTRPG--VDALMT-----TCPVTDGGGVPFSPKRL----- 342
 Qy 394 LTNALPNSPIAGQWPPADQOLVANR 420
 Db 343 ---QLPRKPAAGR---GCRDTIVRSAR 363

RESULT 14

E86719

hypothetical protein yhd (imported) - Lactococcus lactis subsp. lactis (strain IL1403)

C:Species: Lactococcus lactis subsp. lactis

C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001

C:Accession: E86719

R:Botolin, A.; Wincker, P.; Manger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrlich

Genome Res. 11, 731-753, 2001

A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss

A:Reference number: A86625; MUID:21235186; PMID:11337471

A:Accession: E86719

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1180 <STO>

A:Cross-References: GB:AE005176; PID:g12723674; PIDN:AAK04855.1; GSPDB:GN00146

C:Genetics:

A:Gene: yhd

Query Match 5.3%; Score 120; DB 2; Length 1180;

Best Local Similarity 21.4%; Pred. No. 1.6; Mismatches 155; Indels 178; Gaps 26;

Matches 105; Conservative 53; Mismatches 155; Indels 178; Gaps 26;

Qy 6 NPAGATFEVNPYMAOEVOSEANOTNATLAKMRYSTV---STAVMDRIAI--- 57

Db 579 NPVGATSTLNTY---TVTLTAVSQT-----ATYGYTSTALNTPAIPALIS 624

Qy 58 NGVNGRPGILT-----YLD-----ALSQOQT----- 80

Db 625 TGTGG-SVTTPTGYTSPQIPSGYISAIYAGTSASGTLISSSTNGTNMAAPSGOYVA 683

Qy 81 -----TPEVIRIV-----IYDLPRGDCALASNGELPATAGL-----QYEVQYI 121

Db 684 PRGNQYIOLAPRTVNTVFISIDPRNDPLNVAOGGPTPYGSLIGSSTTNQSOYI 743

Qy 122 DPASILSNPKYSSLRITVITIEPDSLPMNAVTNMSIQACATAPVY-----YEOGIEYALT 175

Db 744 DMNNSLTGN-----TLSTTD-----NVTYDGAATGKIYIYWGQYQGGTTTAT 787

Qy 176 KLHAIRNVITYD--AAHSGML-GMPNNAAGYQEVQKVLNMSIGVNGIDGFTNTA-NY 231

Db 788 NSAPITTYKTFSDLAHAYVLAGTSMNRTOLVHLMTVDQTTITSNSTSGHVPATNY 847

Qy 232 TPLKEPFMTATQOVGQPVESANFYQWNPDIADADYAVDLYSRLVAAGPSSIGMLIDL 291

Db 848 IRSSF--ITSSDIDG-----ADDDGLSTAINGNATV----- 878

Qy 292 RKGWGPNEPTGPSTADVTNVTNOSKIDLRQHRGLMCNQNGAGLQPPQASPTDEPNNAH 351

Db 879 -NITGGQTOYWGAT-----NSQLSIGVTENYVALNTYG----- 913

Qy 352 LDAYV-WIRPGES-----DGTSAASDPTTGKKSDDPCDPT--YTSYGVLTNA 397

Db 914 VQAYTLM--PSHSGTQSQFLOTLDTATSLANDVTYSQFTLEVDATATYTNHNIIPRA 971

Qy 398 L--PNSPIAG 405

Db 972 ASINPSSDLVG 982

RESULT 15

D90441

ABC transporter, periplasmic component SSO2669 (imported) - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001

C:Accession: D90441
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Ch
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder
 aretti, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.
 submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: D90441

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-693 <KUR>

A:Cross-References: GB:AE006641; NID:g13815984; PIDN:AAK42787.1; GSPDB:GN00155

C:Genetics:

A:Gene: SSO2669

Query Match 5.0%; Score 113.5; DB 2; Length 693;

Best Local Similarity 22.1%; Pred. No. 2.2; Mismatches 190; Indels 107; Gaps 26;

Matches 102; Conservative 63; Mismatches 190; Indels 107; Gaps 26;

Qy 7 PYGATFEVNPYMAOEVOSE--AANOTNATLAK-----NRVSTSTAVMDRIAIAG 59

Db 242 PYG-----PFLSOLSNHEIYLNKRPYIDAKYIPFNOYIYQGA---DLAA--A 230

Qy 60 VNGG-----PGLTYLDAALSOQO--GTPPEVIRIYDLPGRDCALASNGELPATAG 112

Db 291 ITGGVSMDSMGLT-----GLSPQLLESPLTLEVI--NLPDPFGMGIAFNLPN---W 339

Qy 113 LQYEVQYIDPILASISNPKYSSLRITVITIEPDSLPMNAVTNMSIQACATAPVYEOGIEY 172

Db 340 LREYQVR--AAIAYILNRATIASVG--GLTAPVITIPNAPNLSYSPMTSSOYSSLPY 396

Qy 173 --ALTK-LHAIPNVYIYMDAAHSGMLG-----PNNAGYQEVQKVL 212

Db 397 NVNLTKAQLKSVGYQKSG-----VWTPNCTPPLITIGASPPFQALAMAEVOKEL 451

Qy 213 NASIGN-----GIDGFTN--TANTPPLKEPFMTATQOVGQPYE 251

Db 452 Q-QFGINVOLHIYVVSQWHAQWONGGYDLMFENNGSSVPGTAPWSLVLSFGGP-- 508

Qy 252 SANFYQWNPDIADAD--YAVDLYSRLVAAGPSSIGMLIDLKNGWGPNE---PTGPTSA 307

Db 509 -WNTQWNNENTLPLNGTITIDFHLLEETESPNTQOLIQANQELSYVMNOYLLPILPVE 567

Qy 308 TDVNTFVNOSKIDLRQHRGLMCNQNGA--GLGQPPQASPTDEPNNAHLDAYVWIRPGES-- 364

Db 568 IENYVIVNPSLLTAAPRANSMWEBAQYIG-----GTAMVALID--YWAPLYESTI 619

Qy 365 --DGTSAASDPTTGKKSDDPCDPTYTSYGVLTNALPNSPI 403

Db 620 ITTTTTSISTTITTTTTSATTAATTSVTTTTSVTTTTSISTTV 661

Search completed: August 29, 2002, 16:24:03
 Job time: 338 sec

...

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:38:31 ; Search time 38.44 Seconds
(without alignments)
426.076 Million cell updates/sec

Title: US-09-917-384-4

Perfect score: 2249
Sequence: 1 ATHVDPYAGATFFVNPYMA.....AGQWFPAGDQLVANRPAY 423

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1244	55.3	872	1	GUXX_CELFI
2	639.5	28.4	438	1	GUXX_AGABI
3	558	24.8	462	1	GUXX_FUSOX
4	546	24.3	471	1	GUXX_TIRE
5	261.5	11.6	441	1	GUXX_THERF
6	244.5	10.9	449	1	GUXX_CELFI
7	238.5	10.6	456	1	GUXX_MITCI
8	193.5	8.6	359	1	GUXX_STRSQ
9	181.5	8.1	321	1	GUXX_STRHA
10	112	5.0	790	1	QUXA_XANCT
11	107	4.8	657	1	CN16_HAEIN
12	107	4.8	657	1	GAOA_DACDE
13	103.5	4.6	527	1	VG12_BPT4
14	100.5	4.5	321	1	EL3F_HORVU
15	99	4.4	739	1	GUXX_CLOTM
16	99	4.4	2124	1	PGCA_RAT
17	98.5	4.4	443	1	HXA3_MOUSE
18	98.5	4.4	512	1	YES6_YEAST
19	98.5	4.4	1136	1	CABA_BACTI
20	98.5	4.4	1567	1	PMN2_MOUSE
21	98	4.4	423	1	P65_MYCPN
22	97.5	4.3	561	1	LCFA_ECOLI
23	97	4.3	1005	1	Y456_CHLTR
24	96	4.3	3354	1	CADN_HUMAN
25	96	4.3	910	1	DNJM_MYCPN
26	96	4.3	1196	1	AMTB_PAEPO
27	95.5	4.2	474	1	ENCB_HUMAN
28	95	4.2	433	1	ENO_BORBO
29	95	4.2	2132	1	PGCA_MOUSE
30	94.5	4.2	1589	1	PHP_DROME
31	94.5	4.2	1742	1	GUXX_CALSA
32	94	4.2	561	1	HYVB_PROMI
33	94	4.2	625	1	AMTG_NEUCR

34	94	4.2	768	1	AMT1_SACDI	P29160	saccharomyc
35	94	4.2	936	1	CDAS_HUMAN	Q9517	homo sapien
36	93.5	4.2	294	1	MTAI_SOREQ	O13595	soraria eq
37	93.5	4.2	837	1	LDLR_RABIT	P20063	oryctolagus
38	93	4.1	530	1	HLX_HAI17	P29143	halophilic
39	93	4.1	809	1	COAT_DSDNV	Q71155	diatraea sa
40	93	4.1	937	1	CDAY_HUMAN	Q9UN72	homo sapien
41	93	4.1	1134	1	YML7_YEAST	Q03735	saccharomyc
42	92.5	4.1	357	1	SFAI_STRFR	Q03424	streptomyc
43	92.5	4.1	409	1	ALP_TRIHA	Q03420	trichoderma
44	92.5	4.1	409	1	GUN2_BACSA	P06565	baclillus sp
45	92.5	4.1	950	1	CDAL_HUMAN	Q95113	homo sapien

ALIGNMENTS

```

RESULT 1
GUXX_CELFI          STANDARD;          PRT;          872 AA.
ID      GUXX_CELFI
AC      P50401;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      Exoglucanase A precursor (EC 3.2.1.91) (Exocellulohydrolase A)
DE      (1,4-beta-cellulohydrolase A) (CBP95).
GN      CBHA.
OS      Cellulomonas fimi.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
RX      NCBI_TaxID=1708;
RP      SEQUENCE FROM N.A.
RC      SRRAIN-ATCC 484;
RX      MEDLINE=94344030; PubMed=8065260;
RA      Meinke A., Gilkes N.R., Kwan E., Kilburn D.G., Warren R.A.J.,
RA      Miller R.C., Jr.;
RT      "Cellulohydrolase A (Cbha) from the cellulolytic bacterium
RT      Cellulomonas fimi is a beta-1,4-exocellulohydrolase analogous to
RT      Trichoderma reesei CBH II."
RL      Mol. Microbiol. 12:413-422(1994).
RN      [2].
RP      SEQUENCE OF 41-58.
RX      MEDLINE=93209933; PubMed=8458833;
RA      Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C., Jr., Warren R.A.J.;
RT      "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
RT      D (Cend), a family A beta-1,4-glucanase."
RL      J. Bacteriol. 175:1910-1918(1993).
CC      - FUNCTION: THIS ENZYME HYDROLYSES 1,4-BETA-D-GLUCOSIDIC LINKAGES OF
CC      CELLULOSE. WEAK ACTIVITY AGAINST CARBOXYMETHYLCELLULOSE, BACTERIAL
CC      MICROCRYSTALLINE CELLULOSE AND BARLEY BETA-GLUCAN. HAS ALSO WEAK
CC      ENDOLUCANASE ACTIVITY. HYDROLYSES GLUCOSIDIC BONDS WITH INVERSION
CC      OF ANOMERIC CONFIGURATION.
CC      - CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC      in cellulose and cellotetraose, releasing cellobiose from the non-
CC      reducing ends of the chains.
CC      - SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC      - SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC      (CBD).
CC      - SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL
CC      HYDROLASES).
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
DR      EMBL: L25809; AAC36898.1; -.
DR      HSP: P07986; IEXG.
DR      InterPro: IPR001919; CBD_2.

```

DR InterPro: IPR003961; FN.III.
DR InterPro: IPR003962; FN.III.repeat.
DR InterPro: IPR001524; Glyco_hydro_6.
DR Pfam: PF00553; CBD_2; 1.
DR Pfam: PF00041; fn3; 3.
DR Pfam: PF01341; Glyco_hydro_6; 1.
DR PRINTS: PRO0014; FNTYPEIIT.
DR PRINTS: PRO0733; GLHYDRASE6.
DR PRODOM: PD003733; Glyco_hydro_6; 2.
DR SMART: SM00060; FN3; 3.
DR PROSITE: PS00561; CBD_BACTERIAL; 1.
DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT SIGNAL 1 40
FT CHAIN 41 872 EXOGLUCANASE A.
FT DOMAIN 41 477 CATALYTIC.
FT DOMAIN 478 563 FIBRONECTIN TYPE-III 1.
FT DOMAIN 573 664 FIBRONECTIN TYPE-III 2.
FT DOMAIN 673 768 FIBRONECTIN TYPE-III 3.
FT DOMAIN 769 872 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 188 872 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 410 410 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 140 202 BY SIMILARITY.
FT DISULFID 374 428 BY SIMILARITY.
FT DISULFID 770 869 BY SIMILARITY.
SQ SEQUENCE 872 AA; 89300 MW; 7883BA407E995533B CRC64;

Query Match 55.3%; Score 1244; DB 1; Length 872;
Best Local Similarity 55.7%; Pred. No. 2,6e-82;
Matches 239; Conservative 60; Mismatches 120; Indels 10; Gaps 6;

QY 3 HYDNYAGATFFVNPYMAOEVOSEANOT-NATLAKKRVSTYSTAVMDRIAINGVN 61
DB 44 HVDNYAGATFFVNPYMAOEVOSEANOT-NATLAKKRVSTYSTAVMDRIAINGVN 103
QY 62 GGGPLTYIDALSOOGG-TPREYIEIYTDJGRDCAALASNGELPATAALQYETGY 120
DB 104 DNGKLFHIDNNAVOOKAAGVPLVNLVLYTDJGRDCAALASNGELPATAALQYETGY 163
QY 121 IDPIASILSNPKYSSILRYTIIPEDSLPAVNTMSIQACATVAPYEGIEYALTKLHAI 180
DB 164 IDPIADLDNPEYESIRIATIEPDSLRLTINISEPNOQAAPYRGVAKALDKLHAI 223
QY 181 PNVYIYMDAHSGLGMPNNA---SGYVOEOKVLINASTGVNGIDGFTVNTANTPLKEP 237
DB 224 PNVYIYIDIGHSGWLGWDSNAGPSATLFEVAK--STAGFASIDGFVSDVANTPLEEP 281
QY 228 FMT-ATQOYGQGVESANFYOMNPIDIEADYAVDLYSRVYAGFPSSIGMLDTLRNGW 296
DB 282 LISDSSLITINNTPIRSSKFTYEMNPFDELDYTAHMRRLVLAAGFPSSIGMLDVTSRNGW 341
QY 297 GNEPTGSTATVNTFVNSKIDLRHRLGVLNONGAGLGPPOASPTDFPNHLDAVY 356
DB 342 GNRPTSTIASDVANVADNRVHRGANCNPLGAGIGFEPATEPGYAAASHLDAYV 401
QY 357 WKPGESEDSCTSAADPTTGKSDPCDCPTTYTS--YGVLTALPNSPIAGCWFPQAFQ 414
DB 402 WKPGESEGSASTIDIPNDGKRFDRMCDPTFVSPKLNQLTGATNPANAPLAWFMEQFV 461
QY 415 LVANARPAV 423
DB 462 LVKNAPV 470

RESULT 2
ID GUX3_AGABI STANDARD; PRT; 438 AA.
AC P49075;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Exoglucanase 3 precursor (EC 3.2.1.91) (Exocellobiohydrolase 3)
DE (1,4-beta-cellubiohydrolase 3).
GN CEL3.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 255-277 AND 331-351.
RC STRAIN=D649;
RX MEDLINE=94368092; Pubmed=8085821;
RA Chow C.-M., Yague E., Raguz S., Wood D.A., Thurston C.F.;
RT "The cel3 gene of Agaricus bisporus codes for a modular cellulase and
RT is transcriptionally regulated by the carbon source";
RT Appl. Environ. Microbiol. 60:2779-2785(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Mycelium;
RX MEDLINE=96269930; Pubmed=8662210;
RA Yague E., Chow C.-M., Challen M.P., Thurston C.F.;
RT "Correlation of exons with functional domains and folding regions in
RT a cellulase from Agaricus bisporus";
RL Curr. Genet. 30:56-61(1996).
CC -! FUNCTION: SHOWS ENZYMATIC ACTIVITY TOWARDS CRYSTALLINE CELLULOSE.
CC CELLULOSE AND BARLEY B-GLUCAN.
CC -! CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellobiose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -! SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL
CC HYDROLASES).
CC -! SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: L24519; AAA50607.1; -;
DR EMBL: L24520; AAA50608.1; -;
DR EMBL: Z34007; CAA83971.1; -;
DR HSSP: P00725; I426.
DR InterPro: IPR002524; CBD_fungal.
DR InterPro: IPR001524; Glyco_hydro_6.
DR Pfam: PF00734; CBD_1; 1.
DR Pfam: PF01341; Glyco_hydro_6; 1.
DR PRINTS: PRO0733; GLHYDRASE6.
DR PRODOM: PD001821; CBD_fungal; 1.
DR PRODOM: PD003733; Glyco_hydro_6; 1.
DR SMART: SM00236; fcbd; 1.
DR PROSITE: PS00562; CBD_FUNGAL; 1.
DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 20
FT CHAIN 21 438
FT DOMAIN 21 59 EXOGLUCANASE 3.
FT DOMAIN 60 87 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 88 438 LINKER.
FT ACT_SITE 215 438 CATALYTIC.
FT ACT_SITE 215 438 PROTON DONOR (BY SIMILARITY).
FT DISULFID 28 45 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 39 55 BY SIMILARITY.
FT DISULFID 170 229 BY SIMILARITY.
FT DISULFID 360 407 BY SIMILARITY.
FT VARIANT 133 133 V -> T.
FT VARIANT 152 152 V -> Q.
FT VARIANT 244 244 V -> I.
FT VARIANT 248 248 N -> D.
FT VARIANT 398 398 N -> S.

DT 01-AUG-1988 (Rel. 08, last sequence update)
 DT 01-MAR-2002 (Rel. 41, last annotation update)
 DE Exoglucanase II precursor (EC 3.2.1.91) (Exocellulohydrolase II)
 DE (CBHII) (1,4-beta-cellulohydrolase).
 GN CBH2.
 OS Trichoderma reesei (Hypocrea jecorina).
 CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 CC Hypocreales; Hypocreaceae; Hypocrea.
 CX NCBI_TaxID:51453;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VTT-D-80133;
 RA MEDLINE=87248061; PubMed=3596237;
 RT Teeri T.T., Lehtovaara P., Kauppinen S., Salovuori I., Knowles J.,
 RT "Homologous domains in Trichoderma reesei cellulolytic enzymes: gene
 RT sequence and expression of cellobiohydrolase II.";
 RL Gene 51:43-52(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OM9414 / RUT C-30;
 RA Chen C.M., Gritzali M., Stafford D.W.;
 RT "Nucleotide sequence and deduced primary structure of
 RT cellobiohydrolase II from Trichoderma reesei.";
 RL Biotechnology 5:274-278(1987).
 RN [3]
 RP SEQUENCE OF 25-44.
 RA Fagerstam L.G., Pettersson L.G.;
 RT "The 1,4-beta-glucan cellobiohydrolases of Trichoderma reesei OM
 RT 9414.";
 RL FEBS Lett. 119:97-100(1980).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=90333255; PubMed=2377893;
 RA Rouvinen J., Bergfors T., Teeri T.T., Knowles J.K.C., Jones T.A.;
 RT "Three-dimensional structure of cellobiohydrolase II from Trichoderma
 RT reesei.";
 RL Science 249:380-386(1990).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=97029636; PubMed=8873646;
 RA Koyula A., Reinikainen T., Ruohonen L., Valkeajärvi A.,
 RA Claessens M., Telemann O., Kleywegt G.J., Szaendings M., Rouvinen J.,
 RA Jones T.A., Teeri T.T.;
 RT "The active site of Trichoderma reesei cellobiohydrolase II: the role
 RT of tyrosine 169.";
 RL Protein Eng. 9:691-699(1996).
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
 CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
 CC (2) EXOCELLULOBIHYDROLASES THAT CUT THE DISSACCHARIDE CELLULOSE
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYSE THE CELLULOSE AND OTHER
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellobiose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- MISCELLANEOUS: T. RESEI PRODUCES TWO DIFFERENT
 CC EXOCELLULOBIHYDROLASES. THEY ARE UNIOE IN THAT THEY CAN HYDROLYSE
 CC CRYSTALLINE CELLULOSE IN THE ABSENCE OF ENDOGLUCANASES.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL
 CC HYDROLASES).
 CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
 CC -----
 CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
 CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
 CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
 CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
 CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
 CC ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M16190; AAA34210.1; .

DR EMBL; M55080; AAA72922.1; .
 DR PIR; A26472; A26472.
 DR PIR; A26160; A26160.
 DR PDB; 3CBH; 15-JAN-91.
 DR PDB; 1CB2; 19-MAR-99.
 DR InterPro; IPR000254; CBD_fungal.
 DR InterPro; IPR001524; Glyco_hydro.6.
 DR Pfam; PF00734; CBD_1; 1.
 DR Pfam; PF01341; Glyco_hydro.6; 1.
 DR PRINTS; PR00733; GLHYDRASE6.
 DR ProDom; PD001821; CBD_fungal; 1.
 DR ProDom; PD003733; Glyco_hydro.6; 1.
 DR SMART; SM00236; fcbd; 1.
 DR PROSITE; PS00562; CBD_FUNGAL; 1.
 DR PROSITE; PS00555; GLYCOSYL_HYDROL_F6_1; 1.
 DR PROSITE; PS00556; GLYCOSYL_HYDROL_F6_2; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal;
 KW 3D-structure.
 FT SIGNAL 1 24
 FT CHAIN 25 471
 FT DOMAIN 25 65
 FT DOMAIN 66 106
 FT DOMAIN 107 471
 FT ACT_SITE 199 199
 FT ACT_SITE 245 245
 FT ACT_SITE 425 425
 FT DISULFID 34 51
 FT DISULFID 45 61
 FT DISULFID 200 259
 FT DISULFID 392 439
 FT CARBOHYD 38 38
 FT CARBOHYD 111 111
 FT CARBOHYD 121 121
 FT CARBOHYD 130 130
 FT CARBOHYD 133 133
 FT CARBOHYD 134 134
 FT CARBOHYD 139 139
 FT CARBOHYD 139 139
 FT CARBOHYD 313 313
 FT CARBOHYD 334 334
 FT MUTAGEN 199 199
 FT MUTAGEN 245 245
 FT CONFLICT 359 359
 FT CONFLICT 449 449
 SQ SEQUENCE 471 AA; 49653 MW; C471BC335BDB8 CRC64;
 Query Match 24.3%; Score 546; DB 1; Length 471;
 Best Local Similarity 33.6%; Pred. No. 3e-32;
 Matches 145; Conservative 51; Mismatches 154; Indels 82; Gaps 13;
 QY 1 ATHVDNPVAGATFFVNPVYAOEVOSEANQTNATLAKKRVYSTYAVWMD----- 52
 DB 110 ATYSGRPFVGPVMAWAAVASEVSLAIPSLGAMATAAAVAKVPSFMDLTLPKPLM 169
 QY 53 --RIAMINGVNGPGLITTLDAALSQOOSTPEVIEYIVYDIPGRCAALASNGELPAT 110
 DB 170 EQLTDIRTANNNGG--NAG-----QFVYDLPDRCAALASNGEVSIND 213
 QY 111 AGLQETQETQYIPDIASIIISLPYSSLRVITIEPDSLPVANNMSIQACATVPPYEQGI 170
 DB 214 GGVAKTK-NYIDITIRIV--VEYSDIRTLIVIEPDSLANLVNLTGPKKANNQSAVLECI 270
 QY 171 EYALFKLAIIPNVYIYMDAHSGLGMPNNAAGVVOEVQKVLNASTIGVNGIDGFTYNTAN 230
 DB 271 NYAVTQLN-LPVMVAYLDAGHGAGWPNQDPAQLFANVYKKNSSPRALGLATNTVAN 329
 QY 231 YTPLEKPEFTATQOQGGQVYESANFTYQMPDIDEADYAVDLYSRLVAGFPSSIGMLDT 290
 DB 330 YN-----GWNITSPPSYTGQNAVYNEKLYIHAIGPLLANHGW-SNAEFITDQ 375
 QY 291 LNRNGMGNEPTGPSPATVDTVFVNGSKLDLRHGILWONGAGAGIGQPOASPTDFPRA 350
 DB 376 GRSG-----KQPTG-----QQQWQWQCNVIGTIGFGIRPSANTGD--S 410


```

Db 418 LPEBDS-----ACNGCPAGCGMOEIALEMARNA 447
QY 420 R 420
Db 448 R 448

RESULT 7
GUNA_MICBI STANDARD; PRT; 456 AA.
ID GUNA_MICBI
AC P26414:
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase).
GN CELA.
OS Microbisporea bispora.
OC Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae;
OC Actinomycetales; Streptosporangineae;
OC Unclassified Streptosporangineae; Thermobispora.
OX NCBI_TaxID=2006;
RN [1]
RP SEQUENCE FROM N.A.
RA Yablonsky M.D., Elliston K.O., Eveleigh D.E.;
RL (in) Coughlan M.P. (eds.);
RL Production, characterization and application of cellulose,
RL hemicellulose and lignin enzyme degrading systems, pp.77-83, Elsevier,
RL London (1989).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY B (FAMILY 6 OF GLYCOSYL
CC HYDROLASES).
CC HSSP: P26222; 1TM.
DR Interpro: IPR001919; CBD_2.
DR Interpro: IPR001524; Glyco_hydro_6.
DR Pfam: PF00553; CBD_2; 1.
DR Pfam: PF01341; Glyco_hydro_6; 1.
DR PRINTS: PR00733; GLYHDRLASE6.
DR PRODOM: PD003733; Glyco_hydro_6; 1.
DR PROSITE: PS00361; CBD_BACTERIAL; 1.
DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
DR KEGG: Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT CHAIN 1 30
FT SIGNAL 1 30
FT CHAIN 31 456
FT DOMAIN 31 322
FT DOMAIN 322 335
FT DOMAIN 335 356
FT ACT_SITE 113 113
FT ACT_SITE 113 113
FT ACT_SITE 151 151
FT ACT_SITE 300 300
FT ACT_SITE 300 300
FT DISULFID 114 159
FT DISULFID 267 302
FT DISULFID 360 453
SQ SEQUENCE 456 AA; 47011 MW; B06D8595E322848F CRC64;

Query Match 10.6%; Score 238.5; DB 1; Length 456;
Best Local Similarity 25.4%; Pred. No. 4.7e-10;
Matches 121; Conservative 41; Mismatches 125; Indels 189; Gaps 27;

QY 5 DNYVAGTFVNPYMAQGVSEANQNTATLAKMRVSVSYTAVMMDRIAAT-NG---V 60
Db 36 DSP-----FVDP-----OSNAKVVAA-----PNDPRFVIRDRIAAVPTGWFPA 77
QY 61 NGPG-----LTTYLDALSQOQGTPEVEIYVYDLPGRDCAALANGELPATPAAGIQ 115
Db 78 NYNPSTYRAVDAYVGAAM--AGKIP---IMVYAMPNDCCGSPSAGS-----ARNHT 126
QY 116 YETQYIDPIASISNPKYSSLRITVITIEPDSLPAVATNMS-----IQACATAVPYE 169

```

```

Db 127 AYRAMIDEIAGLRNRP-----AVILLEPDALPTMNCMSSEQAEOVASVGA-----G 176
QY 170 IEYALTKLHAIPNYIYMDAAHSGMLGMPNNAAGVQVQVOKLVASIGVNGDEFTNTA 229
Db 177 KFKFAASSQA-----KYFEDACHDAWVPADEMAS-----RLRAD1-ANSADIALNVS 224
QY 230 NYTPLEKPEMTATQOVGGQPVESANFYQWNPDIADYADVLSRLVAAGPSSIGMLID 289
Db 225 NY-----RYTSGI-----ISVAKSVLAIGA-----SHLRVID 253
QY 290 TLKNGMGCPNEPPTATDVNTVFNOSKIDLRHRLGKLNQNCAGAGCPQASPTDFPN 349
Db 254 TSRNG-----NGPLGE-----WCDPPCRATG---TWSTTPTGD 284
QY 350 AHDYAVYWKPPGSDGTS-----ASDP-----TGKSDP 381
Db 285 PAIDAFIMIKRPGADCIATPGVFVDRAYELAMNAAPPTSPSPPTSPSPSODP 344
QY 382 -----WCDPYTTSYGVLTNLP-----NSPIAG---QW 407
Db 345 GSPSPSPSPAPGRACENTY-----ALVNWPGGFGQAEVTKNKGSSPINKWTYQW 395

RESULT 8
GUNI_STRSQ STANDARD; PRT; 359 AA.
ID GUNI_STRSQ
AC P13933:
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Endoglucanase 1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (cellulase) (Carboxymethyl cellulase) (CMCase I).
GN CASA.
OS Streptomyces sp. (strain KSM-9).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1931;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 71-95.
RX MEDLINE-88313671; PubMed-3410319.
RA Nakai R., Horiouchi S., Beppu T.;
RT "Cloning and nucleotide sequence of a cellulase gene, casa, from an
RT alkalophilic Streptomyces strain.";
RL Gene 65:229-238(1988).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RX MEDLINE-9318417; PubMed-8422992;
RA Damude H.G., Gilkes N.R., Kilburn D.G., Miller R.C. Jr.;
RA Warren R.A.;
RT "Endoglucanase Casa from alkalophilic Streptomyces strain KSM-9 is a
RT typical member of family B of beta-1,4-glucanases.";
RL Gene 123:105-107(1993).
RN [3]
RP SEQUENCE OF 185-222, AND REVISIONS.
RX MEDLINE-92104156; PubMed-1761039;
RA Gilkes N.R., Claessens M., Aebersold R., Hentisat B., Melnke A.,
RA Morrison H.D., Kilburn D.G., Warren R.A.J., Miller R.C. Jr.;
RT "Structural and functional relationships in two families of beta-1,4-
RT glucanases.";
RL Eur. J. Biochem. 202:367-377(1991).
CC -1- FUNCTION: CMCase I PREFERENTIALLY HYDROLYZES CARBOXYMETHYL
CC CELLULOSE (CMC) AND SHOWS A MAXIMUM ACTIVITY AT PH 8.5.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY B (FAMILY 6 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR	EMBL; L03218; AAA26713.1; -	
DR	EMBL; X61008; CAA43330.1; -	
DR	EMBL; M20921; AAA26776.1; ALT_SEQ.	
DR	PIR; J03038; J03038.	
DR	HSSP; P26222; 1TML.	
DR	InterPro: IPR001524; Glyco_hydro_6.	
DR	Pfam: PF01341; Glyco_hydro_6; 1.	
DR	PRINTS; PR00733; GLHYDRASE6.	
DR	ProDom; PD003733; Glyco_hydro_6; 1.	
DR	PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.	
DR	PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.	
KM	Cellulose degradation; Hydrolase; Glycosidase; Signal.	
FT	SIGNAL	1
FT	PROPEP	?
FT	CHAIN	71
FT	ACT_SITE	154
FT	ACT_SITE	192
FT	ACT_SITE	339
FT	DISULFID	155
FO	SEQUENCE	359 AA; 36514 MW; F1300238834E906C CRC64;

QUTX_XANCU STANDARD; PRT; 790 AA.

ID QUTX_XANCU STANDARD; PRT; 790 AA.

AC 09XD78;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Probable quinone dehydrogenase [pyroloquinoline-quinone] (EC 1.1.99.25).

GN QUMA.

OS *Xanthomonas campestris* (pv. *juglandis*).

OC Bacteria; Proteobacteria; gamma subdivision; *Xanthomonas* group;

OC *Xanthomonas*.

OX NCBI_TaxID=44291;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=CS;

RX MEDLINE=20063481; PubMed=10594704;

RA Lee Y.-A., Lo Y.-C., Yu P.-P.;

RT "A gene involved in quinone metabolism is specific to one DNA homology group of *Xanthomonas campestris*."

RL J. Appl. Microbiol. 87:649-658(1999).

CC -1- CATALYTIC ACTIVITY: Quinate + pyroloquinoline-quinone = 5-dehydroquinone + reduced pyroloquinoline-quinone.

CC -1- COFACTOR: PQQ (BY SIMILARITY).

CC -1- PATHWAY: QUINIC ACID CATABOLIC PATHWAY; FIRST STEP. THIS PATHWAY ALLOWS GROWTH OF BACTERIA WITH QUINATE BY ITS CONVERSION TO PROTOCATHECHUATE AND SUBSEQUENT METABOLISM BY THE BETA-KETOADIPATE PATHWAY.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -1- SIMILARITY: BELONGS TO THE BACTERIAL PQQ DEHYDROGENASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AF109471; AAD38453.1; -

DR InterPro: IPR001479; Bac_PQQ.

DR InterPro: IPR002372; Bac_PQQ_repeat.

DR Pfam: PF01011; Bacterial_PQQ_6.

DR PROSITE: PS00363; BACTERIAL_PQQ_1; FALSE_NEG.

DR PROSITE: PS00364; BACTERIAL_PQQ_2; FALSE_NEG.

KW Oxidoreductase; PQQ; Quinate metabolism; Transmembrane.

FT TRANSMEM 22 42 POTENTIAL.

FT TRANSMEM 48 68 POTENTIAL.

FT TRANSMEM 77 94 POTENTIAL.

FT TRANSMEM 106 126 POTENTIAL.

SO SEQUENCE 790 AA; 82896 MW; B75F29B52A49FE6F CRC64;

Query Match 5 0%; Score 112; DB 1; Length 790;

Best Local Similarity 23.3%; Pctd. No. 1.4; Indels 144; Gaps 24;

Matches 107; Conservative 39; Mismatches 170;

QY 43 STYSTAVMRIAINGVN-----GGGLTYTIDALISQOOGT---TP-EVIRIYI 89

DB 152 SAAGNTDGGSRFALDQINSNGRPAAGSPPTTGGRIANSDGAGADQLTLPQVGEKVF 211

QY 90 YDLPGRCDAAL-ASNG-----ELPATVA-----GLQTYETQYIDIASILSPKYSYL 136

DB 212 LCTPHNNLILALDASTGKQLMRREINATTSVMQRCRGIGYFDADALPAPSV-ANP--SPI 268

QY 137 RYITLIERDLPNNAVNTMMSIQACATAPRYEGEGEVALTKLHAIIPNYIYIMDAHSGMLG 196

DB 269 AATYVAGACRCRRRLFTNTIDGRILAV-----DAQTGA-- 302

QY 197 WPNASGYVOEVQKVLNASIGVNGIDGFVNTANTYPLKEPFMTATQOVGQ----- 248

DB 303 -----CGFGSGNGVDLKAIGLA-APDPF-----YQLTSPPLVAGTIVGGRTRADNDVQ 351

QY 249 -----PVSAIFYQNNPDIDEADYAVDLKSLVA-AGPSSIGMLIDTLRN 293

DB 352 TDMPGVYRGSMSPVSSAGIDPCNPHDRQAPAAAGSSVSRSTPVWMA-PMSDAMNTVFL 411

QY 294 GMGGPN-----EPT-----GPST-ATDVNT-----FVNSKIDLRHGRIGMCNQNGAGL 336

DB 412 PLGGPSYDLGAERTALDHRYGASVIALDATTGAKEKWTOT-----VHNDLM-----DF 460

QY 337 GQPPQASPTDFPN--AHLDAYWIKPPGESDGTSAASDPTTGGKSPMCD----- 384

DB 461 DLPMPQSLIDPPNDGSHTRPV-----IGTKAGIIVYLDRTGR--PLTEVREVPYKGS 513

QY 385 -----PYTTSYGVLTNALPNSPIAGOWFPAQFDQLV 416

DB 514 DIAHQYAPTOPLSVGMPOIGTKHLTESDMWGATAMDQML 553

RESULT 11

CN16_HAEIN STANDARD; PRT; 657 AA.

ID CN16_HAEIN STANDARD; PRT; 657 AA.

AC P44764;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE 2',3'-cyclic-nucleotide 2'-phosphodiesterase precursor (EC 3.1.4.16).

GN CPDB OR H10583.

OS *Haemophilus influenzae*.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC *Haemophilus*.

OX NCBI_TaxID=727;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=RD / KW20 / ATCC 51907;

RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M., Greim C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;

RT "Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd."

RL Science 269:496-512(1995).

CC -1- FUNCTION: THIS BIFUNCTIONAL ENZYME CATALYZES TWO CONSECUTIVE REACTIONS CONVERTING 2',3'-CYCLIC-NUCLEOTIDE TO 3'-NUCLEOTIDE AND THEN 3'-NUCLEOTIDE TO NUCLEIC ACID AND PHOSPHATE (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: Nucleoside 2',3'-cyclic phosphate + H(2)O = nucleoside 3'-phosphate.

CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).

CC -1- SIMILARITY: BELONGS TO THE 5'-NUCLEOTIDASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: U33740; AAC22242.1; -

DR TIGR: H10583; -

DR InterPro: IPR002224; 5_nucleotidase.

DR InterPro: IPR000934; Ser_chr_phosphatase.

DR Pfam: PF01009; 5_nucleotidase; 1.

DR Pfam: PF02872; 5_nucleotidase; 1.

DR PROSITE: PS00785; 5_NUCLEOTIDASE_1; 1.

DR PROSITE: PS00786; 5_NUCLEOTIDASE_2; 1.

KW Hydroxylase; Multifunctional enzyme; Periplasmic; Signal; Complete proteome.

FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 657 2',3'-CYCLIC-NUCLEOTIDE 2'-
SEQUENCE 657 AA; 72763 MM; 201CAB415014499 CRC64;
PHOSPHODIESTERASE.
Matches 75; Conservative 33; Mismatches 123; Indels 126; Gaps 16;
Query Match 4.8%; Score 107; DB 1; Length 657;
Best Local Similarity 21.0%; Pred. No. 2.5;
Query 166 YEOGIEYALTKLHAIPNYIYDA--AHSGWL-----GMPNNAS 202
DB 243 YQEGENAFYADVPHT---DAVIFGSHSLFPNKEFASPNADYNGITKGIPESEA 298
QY 203 GT-----VOEVQVYNASISVNGIDGVTNTANTYPIKEPFMTAT 242
DB 299 GYMANNISVVDLGLTEHKGKIVITSGKAVLRPIYDIETKKALAKNDPETALLKPVHEAT 358
QY 243 QOVGQPESEA--NFYQ-----NPDIADYAVDLSRLVAAGFPSSIGMLITLRNG 294
DB 359 KRYVSOPYGRKADNMYSTIALLDQDPTIYVQAQKAYEKVAPSIAAMAGLPI--LSAG 416
QY 295 -----WGGPNEPTGPSTADVTTFYNOSKIDLRHGRGLMCNQN-----G 333
DB 417 APFKAGGRKNDPTG-----YTEVKNKGKLFERNADLVLYPNLTVVYKATGEQLEWLE 469
QY 334 AGLGPPQASPT-DEPNMHLDAIYV-----IKPGESEDGTSASDP 373
DB 470 CSAGMEKQIDPTSDKPSQIDID--WEGFRYVNFYDIDGVNVEYDLTKPARYDGECKLINP 526
QY 374 TT-----GKSDPMDPPTYSYGLVFNALPNSPIAGOWPFAQDQIVANARP 421
DB 527 ESHRYVNLTYQCKPVDPRAE-----FLIATN--NYRAYGKPEPGTGKHIYVASP 574
RESULT 12
GAOA_DACDE
ID GAOA_DACDE STANDARD: PRT: 680 AA.
AC 001745;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Galactose oxidase precursor (EC 1.1.3.9) (GAO).
GN GAO.
OS Dactylium dendroides (Cladobotryum dendroides).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypomyces.
OX NCBI_TaxID=5132;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE-92235025; PubMed-1569070;
RA McPherson M.J., Ogel Z.B., Stevens C.E., Yadav K.D.S., Keen J.N.,
RT Knowles P.F.;
RT "Galactose oxidase of Dactylium dendroides. Gene cloning and sequence
RT analysis";
RL J. Biol. Chem. 267:8146-8152(1992).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RX MEDLINE-91163641; PubMed-2002850;
RA Ito N., Phillips S.E.V., Stevens C.E., Ogel Z.B., McPherson M.J.,
RT Keen J.N., Yadav K.D.S., Knowles P.F.;
RT "Novel thioether bond revealed by a 1.7 A crystal structure of
RT galactose oxidase";
RL Nature 350:87-90(1991).
CC -1- FUNCTION: CATALYSES THE STEROSPECIFIC OXIDATION OF A BROAD RANGE
CC OF PRIMARY ALCOHOL SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: D-galactose + O(2) -> D-galacto-hexodialose +
CC H(2)O(2).
CC -1- COFACTOR: COPPER.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: CONTAINS 3 KELCH REPEATS.
CC -1- DATABASE: NAME-Worthington enzyme manual;

CC CC WWW="http://www.worthington-biochem.com/manual/G/GAO.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M86819; AAL16228.1; -.
DR PDB: 1GOG; 31-JAN-94.
DR PDB: 1GOH; 31-JAN-94.
DR InterPro: IPR000421; FA58_C.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF01344; Kelch. 3.
DR PRINTS: PR00501; KELCHREPEAT.
DR SMART; SM00231; FA58C; 1.
KW Oxidoreductase; Signal; Copper; Repeat; 3D-structure.
FT SIGNAL 1 41
FT CHAIN 42 680 GALACTOSE OXIDASE.
FT REPEAT 279 321 KELCH 1.
FT REPEAT 323 372 KELCH 2.
FT REPEAT 492 544 KELCH 3.
FT DISULFID 59 68
FT THIOETH 269 313
FT DISULFID 556 559
FT METAL 313 313
FT METAL 536 536 COPPER.
FT METAL 537 537 COPPER.
FT METAL 622 622 COPPER.
FT METAL 48 49
FT STRAND 56 59
FT STRAND 63 63
FT STRAND 64 65
FT STRAND 66 66
FT STRAND 68 72
FT TURN 76 77
FT TURN 80 81
FT STRAND 85 87
FT STRAND 94 112
FT TURN 120 121
FT STRAND 123 123
FT STRAND 125 131
FT STRAND 142 145
FT STRAND 153 169
FT STRAND 172 172
FT TURN 173 174
FT STRAND 180 181
FT STRAND 183 188
FT TURN 197 198
FT STRAND 201 207
FT STRAND 214 217
FT TURN 219 221
FT STRAND 224 227
FT TURN 236 237
FT STRAND 243 249
FT TURN 250 253
FT STRAND 254 260
FT TURN 262 263
FT STRAND 271 274
FT TURN 276 277
FT STRAND 279 283
FT TURN 285 286
FT TURN 288 289
FT STRAND 291 295
FT TURN 296 299
FT STRAND 300 303
FT STRAND 312 312
FT STRAND 315 318
FT TURN 320 321
FT STRAND 324 327

FT STRAND 330 330
FT STRAND 332 332
FT STRAND 340 340
FT TURN 345 348
FT STRAND 349 353
FT TURN 354 355
FT HELIX 359 361
FT TURN 366 368
FT HELIX 369 372
FT TURN 373 373
FT STRAND 378 380
FT HELIX 382 384
FT STRAND 386 388
FT STRAND 394 399
FT TURN 402 403
FT STRAND 405 411
FT STRAND 413 414
FT TURN 415 416
FT STRAND 417 418
FT TURN 423 423
FT STRAND 424 424
FT STRAND 426 431
FT TURN 432 435
FT STRAND 436 440
FT STRAND 443 443
FT STRAND 450 450
FT STRAND 452 452
FT STRAND 455 459
FT TURN 463 464
FT STRAND 468 471
FT TURN 473 474
FT STRAND 480 481
FT STRAND 484 487
FT TURN 489 490
FT STRAND 493 496
FT STRAND 499 499
FT STRAND 501 501
FT TURN 504 505
FT STRAND 510 510
FT STRAND 515 518
FT TURN 519 522
FT STRAND 523 526
FT TURN 536 537
FT STRAND 538 542
FT TURN 544 545
FT STRAND 548 552
FT TURN 557 558
FT STRAND 555 570
FT HELIX 572 574
FT STRAND 575 575
FT TURN 577 578
FT STRAND 581 581
FT STRAND 587 590
FT STRAND 594 596
FT TURN 597 598
FT STRAND 600 605
FT STRAND 611 616
FT STRAND 619 621
FT TURN 622 623
FT STRAND 624 624
FT STRAND 630 633
FT STRAND 636 637
FT STRAND 642 646
FT TURN 651 653
FT STRAND 657 664
FT TURN 666 667
FT STRAND 670 670
FT STRAND 674 679
SQ SEQUENCE 680 AA: 72823 MW: 2F97C561B63E46E9 CRC64;

Query Match 4.8%; Score 107; DB 1; Length 680;
Best Local Similarity 21.4%; Pred. No. 2.6;

Matches 109; Conservative 48; Mismatches 156; Indels 196; Gaps 31;
QY 19 MAQEVOS-EAANQTNATLAKMRVSTYSPAVMMDRIAINGVNGP----- 64
D 55 MAVTCDSAGSGNRCN-----KAIDGNKDTFM-----HTFYGANGDKPPHTYTIDMKTT 103
QY 65 ----GLTYIDALASQOQGTTPVEIYIYDLPGRDCALASNGELPATNAGLQYETQY 120
D 104 QNVNGL-----SMLPRDGNQNGMI-----GRHEVYLSDG-----TNW 137
QY 121 IDPIAS-----ILSNPKYSS-----LRVTIIEPDSL-NAVTKMSI-QACATAVPY 166
D 138 GSPVAGSWPFPADSTTKYSNETRPARVRLVALTEANGQPTISAIETWFOASSYTAPOP 197
QY 167 EDCIEYALFKLHAIPNVIYIMDAA---HSG---WLGMPNAGSYQVQKVLNASIGVN 219
D 198 GLGRMGPTIDLPVPA-----AAIEPTSGRVLMSSRYRDAG----- 236
QY 220 GIDGFEVNTANTYTP---LKEPPTATQO-----VGQPVESANFY-- 256
D 237 GSPGGLTMSMDPSTGIYSDRTVTWKHDMFCPSIMDNGQIVVTGDNDAKTSLYDS 296
QY 257 ---QWNPDIDEADYAVDLYSRLVAAGPSSIGM---LIDPLRGMW-GGRNEPTG---PS 305
D 297 SSDSWIPGPD-----MQVARGQSSATMSDGRVFTIGSSWSGVEKNGEYISPS 346
QY 306 TAT-----DVNTFVNQSKIDLRHGLMCNONGACL-----GQPPQASPTDFPNAHLD 353
D 347 SKTWTSLPNAKVPMPLTADK-----QGLYRSNDHMLFQWKKGSVYQAPSTRAM----- 396
QY 354 AYWVIRPPGESDGTSAAS-DPTTGKRSDPYC-----DPTY-----TT 389
D 397 ---WYTTSGSGDVKSAGKROSNRGVAPDAMCGNAVMTDAVKKILFRGSPDQODSDATY 453
QY 390 SYGVLT-----NALPNSPIA--GQPFPAQF 412
D 454 NAHITLGEPTGSPNTVFASNGLYFARTF 482
RESULT 13
VG12_BPT4
ID VG12_BPT4 STANDARD: PRT: 527 AA.
AC P10930;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Short tail fiber protein (Protein Gp12) (p12).
GN 12.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like phages.
OX NCBL_TaxID-10665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D;
RX MEDLINE-88189824; PubMed-3357780;
RA Seliyanov N.A., Piliipov A.G., Mesyanzhinov V.V.;
RT "Nucleotide and deduced amino acid sequence of bacteriophage T4 gene
12.",
RL Nucleic Acids Res. 16:2334-2334(1988).
RN [2]
RP SEQUENCE FROM N.A.
RA Kuter E., Arisaka F., Kunisawa T., Taugita A., Mosig G.,
RA Mesyanzhinov V., Ruger W., Stidham T., Thomas E.;
RT "Bacteriophage T4 genome analysis.",
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-57 FROM N.A.
RC STRAIN-D;
RX MEDLINE-89356257; PubMed-2548819;
RA Barrett B.K., Berget P.B.;
RT "Using transposon Tns insertions to sequence bacteriophage T4 gene
11.",

RA da Silveira J.F., Silvestri M.L.Z., Siqueira M.J., de Souza A.A.,
 RA de Souza A.P., Terezi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zaigo M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"
 RL Nature 406:151-159(2000).
 DR EMBL: AE003960; AAF84076.1; -
 DR HSSP: P07987; 1CB2.
 DR InterPro: IPR001524; Glyco_hydro.6.
 DR Pfam: PF01341; Glyco_hydro.6; 1.
 DR PRINTS: PR00733; GLHYDRASE6.
 DR ProDom: PD003733; Glyco_hydro.6; 2.
 DR Complete proteome.
 KW
 SQ SEQUENCE 683 AA; 70852 MW; DC8FB5F76B8ED5F CRC64;

Query Match 52.1%; Score 1171; DB 16; Length 683;
 Best Local Similarity 53.0%; Pred. No. 2e-75;
 Matches 227; Conservative 59; Mismatches 130; Indels 12; Gaps 6;

DB 21 HVDNPFVADSTYLNPDYSKIDTSLAKVNDVSLKEKMQTKYKPGVWIDTIDAVKGP 80
 QY 3 HVDNPFVADSTYLNPDYSKIDTSLAKVNDVSLKEKMQTKYKPGVWIDTIDAVKGP 80
 DB 63 GFG---LFTYDLAALSOOGCTPEVIEIYIYDLPGRDCAALASNGELPATAAGLQTYET 119
 DB 81 NGRRLNLTGHLDAALAAQKGNTPITATFTYIDLPGRDCAALASNGELPATAAGLQTYET 140
 QY 120 YIDPIASILSNPKYSLRTVITIEPDSLPAVYVNSIQACARA--APVYEGGIEVALTK 177
 DB 141 YIDPIASILSNPKYSLRTVITIEPDSLPAVYVNSIQACARA--APVYEGGIEVALTK 200
 QY 178 HALPNVYIYMDAHSGLWPNPNASGYVOEVQVL-NASIGVNGIDGFTYNTANTYPLKE 236
 DB 201 SEIPVYNTYMDIGHSGMLWMDNRLPAVSLYKTVIOSTAGFASVGFATNTANTYPLKE 260
 QY 237 PFM-TATQVGGQVPEANFYQNPDIADYADVLSRLVYAGFPSSIGMLIDTLRNG 295
 DB 261 PNLPNDLNIGGPIRSSKFTYENRKFEDMDSETLYNPFVAGWSSIGFTIDTLRNG 320
 QY 296 GGPNEPTGSTATDVNTFVYVNSKIDLRQHRGLMGNONGAGLQPPQASPTDPNNAHLDA 355
 DB 321 GGPNEPTGSTATDVNTFVYVNSKIDLRQHRGLMGNONGAGLQPPQASPTDPNNAHLDA 375
 QY 356 VWIKPPGESDGTSAASDPTTGKSDPMCDPTTTSIGVLTNLPNSPIAGWFPAPDOL 415
 DB 376 VWIKPPGESDGTSAASDPTTGKSDPMCDPTTTSIGVLTNLPNSPIAGWFPAPDOL 435
 QY 416 VANARPAV 423
 DB 436 INNAIYFDI 443
 RESULT 2
 ID 086730 PRELIMINARY; PRT: 579 AA.
 AC 086730;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE PUTATIVE SECRETED CELLULOSE.
 GN SC567.33.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxId-1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Seeger K.U., Harris D.;
 RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
 RP [2]
 SEQUENCE FROM N.A.

RC STRAIN-A3(2);
 RA Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE-97000351; PubMed-8843436;
 RA Redenbach M., Kleser H.M., Denepalte D., Elchner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb *Streptomyces coelicolor* A3(2) chromosome,"
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL031515; CA20645.1; -
 DR HSSP: P07986; 1EXG.
 DR InterPro: IPR001919; CBD_2.
 DR InterPro: IPR001524; Glyco_hydro.6.
 DR Pfam: PF00553; CBD_2; 1.
 DR Pfam: PF01341; Glyco_hydro.6; 1.
 DR PRINTS: PR00733; GLHYDRASE6.
 DR ProDom: PD003733; Glyco_hydro.6; 2.
 DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
 DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
 SQ SEQUENCE 579 AA; 60831 MW; 5A51DD28E796A89 CRC64;

Query Match 46.7%; Score 1050; DB 2; Length 579;
 Best Local Similarity 49.8%; Pred. No. 6.9e-67;
 Matches 214; Conservative 61; Mismatches 125; Indels 30; Gaps 8;

QY 4 VDNPFAGATFFVNPYVQVSEANQNTNATLAARKRVSTYAVMMRIAINGVNG 63
 DB 165 VDNPFAGATFFVNPYVQVSEANQNTNATLAARKRVSTYAVMMRIAINGVNG 214
 QY 64 PGLITTYDLAALSOOGCTPEVIEIYIYDLPGRDCAALASNGELPATAAGLQTYET 123
 DB 215 MGLRDHIDAL-EQKSGEMVYQVLYNLPGRDCAALASNGELPTE-IDRYKTEYIDP 271
 QY 124 IASISNPKYSLRTVITIEPDSLPAVYVNSIQACAT-----AVPYEGGIEVALTK 176
 DB 272 IASISNPKYSLRTVITIEPDSLPAVYVNSIQACAT-----AVPYEGGIEVALTK 331
 QY 177 LHAIPNVYIYMDAHSGLWPNPNASGYVOEVQVLNASIG--VNGIDGFTYNTANTYPL 234
 DB 332 LGAVGVVYVYVADAGHGMWLGMDSN-RGPSAEIFKTAATTEGATIDYHFTYNTANTYPL 390
 QY 235 KEPEMTATQVGGQVPEANFYQNPDIADYADVLSRLVYAGFPSSIGMLIDTLRNG 294
 DB 391 KEPEMTATQVGGQVPEANFYQNPDIADYADVLSRLVYAGFPSSIGMLIDTLRNG 450
 QY 295 WGPNEPTGSTATDVNTFVYVNSKIDLRQHRGLMGNONGAGLQPPQASPTDPNNAHLDA 354
 DB 451 WGPNEPTGSTATDVNTFVYVNSKIDLRQHRGLMGNONGAGLQPPQASPTDPNNAHLDA 506
 QY 355 YVWIKPPGESDGTSAASDPTTGKSDPMCDPTTTSIGVLTNLPNSPIAGWFPAPDOL 411
 DB 507 YVWIKPPGESDGTSAASDPTTGKSDPMCDPTTTSIGVLTNLPNSPIAGWFPAPDOL 566
 QY 412 PDOLVANARP 421
 DB 567 FOELMKNAV 576
 RESULT 3
 ID 09KH72 PRELIMINARY; PRT: 454 AA.
 AC 09KH72;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE CELLULOSE PRECURSOR.
 GN EX.
 OS Thermomonospora fusca.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptosporangineae; Thermomonosporaceae;

OC Thermobifida.
 OX NCBI_TaxID=2021;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA A.Y.-C., Wilson D.B.;
 RT "Genomic DNA sequence encoding Thermomonospora fusca cellulase with
 overexpression";
 RT Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AF268074; AAF75786.1; -
 DR HSSP; P07987; 1CB2; Glyco_hydro.6.
 DR InterPro: IPR001524; Glyco_hydro.6.
 DR Pfam: PF01341; GLHYDRLASE.
 DR PRINTS: PR003733; GLHYDRLASE.
 DR ProDom: PD003733; Glyco_hydro.6; 2.
 DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
 DR Signal.
 KW SIGNAL.
 FT CHAIN 32 454 POTENTIAL.
 FT SIGNAL 32 454 CELLULOSE.
 SQ SEQUENCE 454 AA; 48624 MW; F8CD1C6DA807134 CRC64;

Query Match

44.4%; Score 999; DB 2; Length 454;
 Best Local Similarity 48.3%; Pred. No. 2,2e-63;
 Matches 209; Conservative 55; Mismatches 137; Indels 32; Gaps 9;

OY 4 VDNPYAGATFEVNPYMAQVQSEANQNTATLAKMRVSTYSTAVWMDRIALING---- 59
 DB 37 VDNPEGAKLYVNPWMAKAAEPGGS-----AVANESAVWLDRIAGLEGNDSP 86
 OY 60 VNGSGELTYTLDALASQOQGTTPVEIETIYDLPGRDCALASNGELPATAGLOTETQ 119
 DB 87 TTSGMGLRHLLEAV-RQSGGDPITQIVYIYMLPGRDCAALASNGELPDE--LDRIKSE 143
 OY 120 YIDPIASISLN-PKYSLSRTVITIEPDSLPAVNTMSIOACATAVPY-----YEGIEY 172
 DB 144 YIDPIDIMWDADENLRIVALEIDSLPNTVTVNGSGELCAVMKONGYVNGVGY 203
 OY 173 ALTKLAIIPVYIYMDAHSGLGMPNNAAGYVOEVQKVLAS-IGVNCIDGFTVNTANY 231
 DB 204 ALTKLGEIPVYIYMDAHSGLGMPNNAAGYVOEVQKVLAS-IGVNCIDGFTVNTANY 263
 OY 232 TPKEPMTATQOYGGQPVESANFYQWNPDIADYAVDLSRLVYAGPSSIGMLIDTL 291
 DB 264 SATVEPRLDVNGTVNGQLIKOSKRWDMNOYVDELSTFVQDLRQALIAKGRSDIGMLIDTS 323
 OY 292 RNMGGPNEPTGPSTADVTNTFYVQSKIDLRHRLGLMNCNGAGLGEPVNPNA-----PG 379
 DB 324 RNMGGPNEPTGPSTADVTNTFYVQSKIDLRHRLGLMNCNGAGLGEPVNPNA-----PG 379
 OY 352 LDAYVWIKRPGESDGTSAASDPTTGKSDPMCDPTTYSYGLTN--ALPNSPIAGQWF 408
 DB 380 VDAYVWIKRPGESDGTSAASDPTTGKSDPMCDPTTYSYGLTN--ALPNSPIAGQWF 439
 OY 409 PAQFDOLVANARP 421
 DB 440 SAQFRELANARP 452

RESULT 4
 ID 060029 PRELIMINARY; PRT; 596 AA.
 AC 060029;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE BETA-1,4-EXOCYDOLYLASE PRECURSOR (EC 3.2.1.91).
 GN E3.
 OS Thermomonospora fusca.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptosporangineae; Thermomonosporaceae;
 OC Thermobifida.
 OX NCBI_TaxID=2021;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-YX;
 RA Zhang S., Lao G., Wilson D.B.;
 RT "Characterization of a Thermomonospora fusca exocellulase";
 RL Biochemistry 34:0-0(1995).
 DR EMBL: U18978; AAA62211.1; -
 DR HSSP; P07986; 1EXG.
 DR InterPro: IPR001919; CHD2.
 DR Pfam: PF00553; CHD2; 1.
 DR Pfam: PF01341; Glyco_hydro.6; 1.
 DR PRINTS: PR003733; GLHYDRLASE.
 DR ProDom: PD003733; Glyco_hydro.6; 2.
 DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
 DR Signal; Hydrolase; Glycosidase.
 KW SIGNAL.
 FT CHAIN 39 596 POTENTIAL.
 FT SIGNAL 39 596 BETA-1,4-EXOCYDOLYLASE.
 SQ SEQUENCE 596 AA; 63547 MW; B0FA5277FE7721E0 CRC64;

Query Match

44.4%; Score 999; DB 2; Length 596;
 Best Local Similarity 48.3%; Pred. No. 3,2e-63;
 Matches 209; Conservative 55; Mismatches 137; Indels 32; Gaps 9;

OY 4 VDNPYAGATFEVNPYMAQVQSEANQNTATLAKMRVSTYSTAVWMDRIALING---- 59
 DB 179 VDNPEGAKLYVNPWMAKAAEPGGS-----AVANESAVWLDRIAGLEGNDSP 228
 OY 60 VNGSGELTYTLDALASQOQGTTPVEIETIYDLPGRDCALASNGELPATAGLOTETQ 119
 DB 229 TTSGMGLRHLLEAV-RQSGGDPITQIVYIYMLPGRDCAALASNGELPDE--LDRIKSE 285
 OY 120 YIDPIASISLN-PKYSLSRTVITIEPDSLPAVNTMSIOACATAVPY-----YEGIEY 172
 DB 286 YIDPIDIMWDADENLRIVALEIDSLPNTVTVNGSGELCAVMKONGYVNGVGY 345
 OY 173 ALTKLAIIPVYIYMDAHSGLGMPNNAAGYVOEVQKVLAS-IGVNCIDGFTVNTANY 231
 DB 346 ALTKLGEIPVYIYMDAHSGLGMPNNAAGYVOEVQKVLAS-IGVNCIDGFTVNTANY 405
 OY 232 TPKEPMTATQOYGGQPVESANFYQWNPDIADYAVDLSRLVYAGPSSIGMLIDTL 291
 DB 406 SATVEPRLDVNGTVNGQLIKOSKRWDMNOYVDELSTFVQDLRQALIAKGRSDIGMLIDTS 465
 OY 292 RNMGGPNEPTGPSTADVTNTFYVQSKIDLRHRLGLMNCNGAGLGEPVNPNA-----PG 521
 DB 466 RNMGGPNEPTGPSTADVTNTFYVQSKIDLRHRLGLMNCNGAGLGEPVNPNA-----PG 521
 OY 352 LDAYVWIKRPGESDGTSAASDPTTGKSDPMCDPTTYSYGLTN--ALPNSPIAGQWF 408
 DB 522 VDAYVWIKRPGESDGTSAASDPTTGKSDPMCDPTTYSYGLTN--ALPNSPIAGQWF 581
 OY 409 PAQFDOLVANARP 421
 DB 582 SAQFRELANARP 594

RESULT 5
 ID 096V97 PRELIMINARY; PRT; 490 AA.
 AC 096V97;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE CELLOBIOHYDROLASE II-LIKE CELLULOSE CELL.
 GN CEL1.
 OS Orpiniomyces sp. PC-2.
 OC Eukaryota; Fungi; Chytridiomycota; Neocallimastixales;
 OC Neocallimastixaceae; Orpiniomyces.
 OX NCBI_TaxID=50059;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PC-2;

RX MEDLINE=21405733; PubMed=11514516;
 RA Steenbakkers P.J.M., Li X.L., Ximenes E.A., Arts J.G., Chen H.,
 RA Ljungdahl L.G., Op Den Camp H.J.M.;
 RT "Noncatalytic docking domains of celulosomes of anaerobic fungi";
 RL J. Bacteriol. 183:5325-5333(2001).
 DR EMBL; AF177205; AAL01212.1; -
 KM Hydrolase.
 SQ SEQUENCE 490 AA; 54051 MW; 85F40D1280759886 CRC64;

Query Match 31.3%; Score 704; DB 3; Length 490;
 Best Local Similarity 39.2%; Pred. No. 2.9e-42;
 Matches 166; Conservative 54; Mismatches 143; Indels 60; Gaps 12;

QY 2 THVDNPYATFEFVNPYMAOEVOSEANOTNATLAARKRVYSTSTAVMMDRIAINGVNG 61
 DB 123 THSGNFEFSVEFFLNPIYVAEVDAAIEOMTNSLKAKAEKMTYSNAIWDITK-----N 177
 QY 62 GSPGLTLYDAALSOOGGTPPEYIEI-VIYDLPGRDCALASNGELPATAGLOTETQY 120
 DB 178 MNGWLETNKAGLQOQEFEGSKVLYFVYDLPGRDCHALASNGELLANDSDAQRIKTEY 237
 QY 121 IDPLASILSNPKYSLRIYITIEPDSLPAVNTNM-SIQACATAVPYEOGIEYALTKLHA 179
 DB 238 IDVIEEKLK--YKKSQPVVLIPEDSLMLVTNLDTSPACRDEKYYLDGHAYLIQKLG 295
 QY 180 INNYIYMAHSGWLGPNNAASGYVOEOKYLNASIGVNG-IDGFVTYNTANTPLKEPP 238
 DB 296 LPHVAATLIDIGHAFWLGMDNREKAKYYSKYI--SSGPKVKRGFTDNNVANTPWEDEPT 353
 QY 239 KPAATQOVGQPVESANFYQWNPDIIDEADYAVDLSRLVAAFPSSIGMLIDTLRNGMG 298
 DB 354 LRGPEPT-----EMNCPDEKRYLEAMHKDFKAAGIESYFFCOTSRNG----- 398
 QY 299 NEPTGPSTATDVNTFVNSKIDLRHRLGIMCNGAGLQPPQASPTDPNNAHLDAIYVNI 358
 DB 399 -----HKYD-RKHPEMCNQTGVGARGPASPVSQMD-YLDAFYNI 438
 QY 359 KPPGESDGTSAADPTTGKSKSPMCDPTTTSYGLVTNALPNSPIAGWPPAFQDOLVAN 418
 DB 439 KPLGESDGT-----DTSARIDGYCGH-----DTAMKPAPEAGWFOKHFEQLENN 485
 QY 419 ARP 421
 DB 486 ANP 488

RESULT 6
 ID 096V98 PRELIMINARY; PRT; 491 AA.
 AC 096V98:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CELLOBIOHYDROLASE II-LIKE CELLULOSE CELH.
 GN CELH.
 OS Orpionmyces sp. PC-2.
 OC Eukaryota; Fungi; Chytridiomycota; Neocallimastcales;
 OC Neocallimastacaceae; Orpionmyces.
 OC NCBI_TaxID=50059;
 RX NCBI_TaxID=50059;
 RP SEQUENCE FROM N.A.
 RC STRAIN=PC-2;
 RX MEDLINE=21405733; PubMed=11514516;
 RA Steenbakkers P.J.M., Li X.L., Ximenes E.A., Arts J.G., Chen H.,
 RA Ljungdahl L.G., Op Den Camp H.J.M.;
 RT "Noncatalytic docking domains of celulosomes of anaerobic fungi";
 RL J. Bacteriol. 183:5325-5333(2001).
 DR EMBL; AF177204; AAL01211.1; -
 KM Hydrolase.
 SQ SEQUENCE 491 AA; 53956 MW; 7C7ACCC9BFF946C4 CRC64;

Query Match 30.8%; Score 692; DB 3; Length 491;
 Best Local Similarity 39.0%; Pred. No. 2.1e-41;
 Matches 163; Conservative 53; Mismatches 142; Indels 60; Gaps 12;

QY 6 NPYAGATFEFVNPYMAOEVOSEANOTNATLAARKRVYSTSTAVMMDRIAINGVNG 65
 DB 130 NPFSGVFFLNPIYVAEVDAAIOMNSLSLKAKAEKMTYSNAIWDITK-----NMQON 184
 QY 66 LTTYIDALSOOGGTPPEYIEIYIDLPGRDCAALASNGELPATAGLOTETQYIDPIA 125
 DB 185 LETNLGALAAQOQSGSKVLYFVYDLPGRDCHALASNGELLANDSDAQRIKTEYIDYIE 244
 QY 126 SILSNPKYSLRIYITIEPDSLPAVNTNM-SIQACATAVPYEOGIEYALTKLHAIPNY 184
 DB 245 EKLR--YKKSQPVVLIPEDSLMLVTNLDTSPACRDEKYYLDGHAYLIQKLGVPVHA 302
 QY 185 IYMDAHSGLWLGPNNAASGYVOEOKYLNASIGVNG-IDGFVTYNTANTPLKEPPATAG 243
 DB 303 MYLIDIGHAFWLGMDNREKAKYYSKYI--SSGPKVKRGFTDNNVANTPWEDEPTLSKRP 360
 QY 244 QVGGQPVESANFYQWNPDIIDEADYAVDLSRLVAAFPSSIGMLIDTLRNGMGPNPPTG 303
 DB 361 DT-----EMNCPDEKRYLEAMHKDFKAAGI-SSVYFVSDTSRNG----- 399
 QY 304 PSTATDVNTFVNSKIDLRHRLGIMCNGAGLQPPQASPTDPNNAHLDAIYVNI 363
 DB 400 -----HKYD-RKHPEMCNQTGVGARGPASPVSQMD-YLDAFYNI 444
 QY 364 SDGTSASDPTTGKSKSPMCDPTTTSYGLVTNALPNSPIAGWPPAFQDOLVAN 421
 DB 445 SDGTS-----DTSARIDGYCGH-----DTAMKPAPEAGWFOKHFEQLENN 489

RESULT 7
 ID 09S3V4 PRELIMINARY; PRT; 351 AA.
 AC 09S3V4:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE 1,4-BETA-CELLULOBIHYDROLASE A (EC 3.2.1.91) (FRAGMENT).
 GN CELA.
 OS Cellulomonas flavigena.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Micrococccineae; Cellulomonadaceae; Cellulomonas.
 OC NCBI_TaxID=1711;
 RX NCBI_TaxID=1711;
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDBB531;
 RA Herrera A., Gutierrez A., Salgado L.M., Ponce-Noyola T.;
 RT "Molecular characterization of cellulases from Cellulomonas
 flavigena";
 RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AF172344; AAD48493.1; -
 DR HSSP; P07987; ICR2.
 DR InterPro; IPR001524; Glyco_hydro.6.
 DR Pfam; PF01341; Glyco_hydro.6; 1.
 DR PRINTS; PR00733; GLHYDRLASE6.
 DR PRODOM; PD003733; Glyco_hydro.6; 1.
 DR PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
 DR Hydrolase; Glycosidase.
 FT NON_TER 351
 SQ SEQUENCE 351 AA; 37615 MW; BCDE63B65DF335C2 CRC64;

Query Match 30.7%; Score 690.5; DB 2; Length 351;
 Best Local Similarity 49.5%; Pred. No. 1.7e-41;
 Matches 151; Conservative 38; Mismatches 93; Indels 23; Gaps 9;

QY 4 VDNPYAGATFEFVNPYMAOEVOSEANOTNATLAARKRVYSTSTAVMMDRIAINGVNG 62
 DB 44 VDNPYAGAVGYVNPYMASSSVRASAKSDASIAAKMGYAKQPTAVMMDRISAITGNAD 103

[illegible]

RESULT	ID	PRELIMINARY:	PRT:	460 AA.
002321	002321			
AC	002321:			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	EXOCYTOLOHIDROLASE.			
GN	CHH1.			
OS	Phanerochaete chrysosporium.			
OC	Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;			
OC	Asphyliophorales; corticiaceae; Phanerochaete.			
OX	NCBI_taxid=5306;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=95110131; Pubmed=7811079;			
RA	Tempeelaars C.A., Birch P.R., Sims P.F., Broda P.;			
RT	"Isolation, characterization, and analysis of the			
RL	chh1 gene of Phanerochaete chrysosporium.";			
RT	Appl. Environ. Microbiol. 60:4387-4393(1994).			
DR	EMBL: S76141; AACB32942.1; -.			
DR	HSSP: P00725; 2CBH.			
DR	InterPro: IPR000254; CBD_fungal.			
DR	InterPro: IPR001524; Glyco_hydro.6.			
DR	Pfam: PF00734; CBD_1; 1.			
DR	Pfam: PF01341; Glyco_hydro.6; 1.			
DR	PRINTS: PR00733; GLHYDRLASE6.			
DR	ProDom: PD001821; CBD_fungal; 1.			
DR	ProDom: PD003733; Glyco_hydro.6; 1.			
DR	SMART: SM00236; fCBD; 1.			
DR	PROSITE: PS00562; CBD_FUNGAL; 1.			
DR	PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.			
DR	PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.			
DR	Hydrobase.			
SO	SEQUENCE 460 AA; 48444 MW; ABBA8A40615FE56B CRC64;			

Query Match	27.5%;	Score 618.5;	DB 3;	Length 460;
Best Local Similarity	36.8%;	Pred. No. 3.4e-36;		
Matches 154;	Conservative 55;	Mismatches 144;	Indels 65;	Gaps 12

```

Oy      5  DNPAGAAFEENPVPMAEVOSEEAANNQATNATLAAMRVSYSTAVVMMDDRTAAI.NGVNGCP 64
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      105 NNPMPTGFQIFLSPYYANEVAAAAAQITDPTLSSKAASVAMIPFTWMDSVAKI-----P 158

Oy      65 GLTYITAAASQOQGT--TPENIEIVYIDDEGRCCAAASGELPATPAAGQIETETQYIDP 123
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      159 DLGTYLASASAKSGTGTQKLVQVQIYITDDEPRCCAAASGSESIANNNGAME--NTID 217

Oy      124 TASTLSNPKYSRLRIVTITIEPDSLPAANAVTNMSIOACATAVPYVEQGLEVALTKLHAIPNV 183
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      218 IVAIQI-QGFDEVRVAIVIEPDSLANTVITNLINOVKCMANATTYLACVNYALMLNAV--GV 274

```

```

QY 184 YIIMDAHSSGLMSPNNASGVYDGEQVYLNASIGVNIIDGVNTATYTLKEPFMATQ 24.3
Db 275 YIMIDAGHAGMLGMPANLSPPAQLFTYQWONAKSPFIKGLATIVANV-----NALQ 32.6
QY 244 QVGGQPVESANFYWMNPIDIEADYAVDLYSLTVAAGFPSSIGMLIDTLRNGMGSPNEPTG 30.3
Db 327 AASPPII-----QGNPIYDEIHNMLAPLLOQAGMDAT--FIVDGRSG----- 37.0
QY 304 PSTATDVTFVYNQSKIDLRQHRGLMCNONGAGLQCPQASPTFDPNAHLDAYVWKPRGE 36.5
Db 371 -----VONIRQMGQMCNKKAGGEGFTRP---TTNNGSOFISDIYVWKPRGE 41.3
QY 364 SDGTSASDPTTGKRSKDMCPRTTTSVGLTNLMLSPSLIAGQFPAQFQVLVANARP 42.1
Db 414 CDGTSNSSP-----RVDSTCS-----LPDAAGPAPEEGTWFQAYFQTLVSAANP 45.8

```

RESULT	9
096V02	
ID	096V02
AC	096V02;
DT	01-DEC-2001 (TREMBLrel. 19, Created)
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE	01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE	CELLULASE CEL6B.
CS	CEL6B.
ON	Lentinula edodes (Shiitake mushroom) (Lentinus edodes).
OC	Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC	Agaricales; Tricholomataceae; Lentinula.
OX	NCBI_Taxid=5353;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=STAMETS CS-2;
RA	Lee C.C., Wong D.W.S., Robertson G.H.;
RT	"Cloning and characterization of two cellulase genes from Lentinula
RL	edodes.";
SD	Submitted (AUG-2001) to the EMBL/genbank/DBJ databases.
EMBL	AF411251; AKK95564.1; -
QO	SEQUENCE 444 AA; 46368 MW; A08BEDD0835FE47C5 CRC64;

Query Match	26.6%;	Score 597.5;	DB 3;	Length 444;
Best Local Similarity	35.3%;	Pred. No. 1e-34;		
Matches 147;	Conservative 60;	Mismatches 143;	Indels 67;	Gaps 14

QY	6	NPVGAFFPVPWPAOEVSAAQNTAATLAAKRRVSTSTAVMMRIAMINGVNGPG	65
Db	92	NPFTGEIYELTSPYANETAAVTAQVTOISDPTLAAAARAKANIPFTFWLQVAKV	145
QY	66	LTFTL-DAAALSQOOGFTTPEVEIEIVTDLPGHDCAAALSNGELPATPAAGLQTYETQYIDPI	124
Db	146	LGTIADLADSAQKBEKKNYLVQIVYVDLPDRDCALASNGEFTIADNGEANYH-DYIDQI	204
QY	125	ASLISNPYSSRLIVTIIIEPDSLBNATNMSIOACATAVPYEEOGIEYALTKLHAIPNVY	184
Db	205	VAQIK--QYPUVHVAAVTEPDSLANTLVNLSVAKCANAQTYIECVYAMQOLSAV-GVY	261
QY	185	IYMAAHSGWLGWPNNSAGVYQVEQKVLNASIGVNGIDGEPYTNANTPLKEPMTATQ	244
Db	262	MYLDAGHAGWGPANLSPAAQLTTSLSYNSGSPSGRGATLVANVNAL-----VATTP	316
QY	245	VGGQPVESANFYQNNPDIIDEADYAVVDYLSRLVAAQFPSSIGMLDTRLNCGMGPNEPTGP	304
Db	317	---DPIT-----QGDPRNDEMLYIEALAPLL--GSFPAH--FIYDQGRSG-----	354
QY	305	STATDVNTFVNQSKIDLROHRGLMCNONGAGLGQPPQASPTDEPNHAIDLVMKIPGES	364
Db	355	-----VQDIRQMGDMCNVLGAGFGPQ---TTNGSSSLIDSIIVWVKPGGEC	398
QY	365	DGTSAADPTTGKKSDDPCDPTTSTTSGVLTNLPSNPJAGOWFPAQFDOLVANARP	421
Db	399	DGTNTSTSP---FYDAHCG-----LPDAPNAPPEAGTGFQOAFETFLVKEAN	442

```
RESULT 10
09C1R4 PRELIMINARY; PRT: 443 AA.
AC 09C1R4;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELLOBIOHYDROLASE.
GN CBH11-1.
OS Lentinula edodes (Shiitake mushroom) (Lentinus edodes).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Tricholomataceae; Lentinula.
OX NCBI_TaxID=5353;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L54;
RA Zhao J., Deane E.E., Kwan H.S.;
RT "Molecular cloning, characterization, and differential expression of a
RL cellobiohydrolase gene from the edible mushroom Lentinula edodes.";
DR EMBL; AF244369; AAK28357.1; -.
DR HSSP; P00725; 1A26.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR001524; Glyco_hydro_6.
DR Pfam; PF00734; CBD_1; 1.
DR Pfam; PF01341; Glyco_hydro_6; 1.
DR PRINTS; PR00733; GLHYDRASE6.
DR ProDom; PD001821; CBD_fungal; 1.
DR ProDom; PD003733; Glyco_hydro_6; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
DR HydroLase.
KW HydroLase.
SQ SEQUENCE 443 AA; 46290 MM; EA9B578BF79EC25A CRC64;

Query Match 26.1%; Score 587.5; DB 3; Length 443;
Best Local Similarity 35.0%; Pred. No. 5.3e-34;
Matches 149; Conservative 60; Mismatches 146; Indels 71; Gaps 15;

OY 1 ATHVNPYAGATF---FVNPMYAOEVOSEANOTNATLAAMRVYSTSTAVMMDR1AA 56
DB 82 STGTTTPAAGNFTQIILSPYANELIAAVTQISDPTAAAKAVANIPFTIMLDQYAK 141
OY 57 INGVNGGGLTYYL-DAALSOQOGTTPREIVIEIVYDLPGRDCAALASNGELPATAG1QT 115
DB 142 V-----PDLGTYLADASAKQSEKKNLYLVYVDLPDRDCAALASNGEFTIADNGEAN 195
OY 116 YETQYIDPIASLSPKYSRLRTVITIEPDSLPAVAVTNMSIQACATAVPYEGGIEYALT 175
DB 196 YH-DYIDIDIVAKIK--QYPDVHVAVVIEPDSLAVNLVTNLSVAKANACQTYTLECYVAAQ 252
OY 176 KLHAIPNYIYIWDAAHSGMLGMPNNAAGVQGVOKVLAASIGVNGIDGVTNTAVTTPAK 235
DB 253 QLSAAV-GVTMTYDAGHAGMLGMPNAPLSPAAQLFTSLYSNAGSPSGVGRGLATVAVYANAL- 310
OY 236 EEFMTATQOVGQPVESANFYQWNPDIIDEADYAVDLYSLVAAGPSSIGMLIDTLRNGM 295
DB 311 ----VAITP---DPIT-----QGPNDYEMLYIEALAPLL--GSPPAH--FTVDGRSG- 353
OY 296 GGPNEPTEPSTATDVNTVFNOSKIDLRQHRGLKCNQNGAGLGQPPQASPTDFPNAHLDAY 355
DB 354 -----VDIRQWGMCMVNLGAGFGTQP---TTNTGSSLIISI 388
OY 356 VWIKPGEISDGTSAASDPTTGKSPKMDPYTTTSGVLTNALPMSPIAGOWFPQAFQDL 415
DB 389 VWVKPGGCDGTSTSSP-----RYDHAG-----LPDATPNAPEAGTWFQAFYETL 435
OY 416 VANARP 421
| | |
```

```
DB 436 VEKANP 441

RESULT 11
09P8N1 PRELIMINARY; PRT: 436 AA.
AC 09P8N1;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELLOBIOHYDROLASE.
OS Coriolus versicolor.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Aphyllophorales; Coriolaceae; Coriolus.
OX NCBI_TaxID=57466;
RN [1]
RP SEQUENCE FROM N.A.
RA Novo C., Clemente A., Simoes F., Mendonca D., Matos J.;
RT "Coriolus versicolor cellobiohydrolase-like cDNA sequence.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF233583; AAF35251.1; -.
DR HSSP; P00725; 2CBH.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR001524; Glyco_hydro_6.
DR Pfam; PF00734; CBD_1; 1.
DR Pfam; PF01341; Glyco_hydro_6; 1.
DR PRINTS; PR00733; GLHYDRASE6.
DR ProDom; PD001821; CBD_fungal; 1.
DR ProDom; PD003733; Glyco_hydro_6; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
DR HydroLase.
KW HydroLase.
SQ SEQUENCE 436 AA; 46170 MM; EDECC5AFBCA324D4 CRC64;

Query Match 25.5%; Score 574.5; DB 3; Length 436;
Best Local Similarity 33.5%; Pred. No. 4.4e-33;
Matches 142; Conservative 65; Mismatches 138; Indels 79; Gaps 14;

OY 6 NPYAGATFVNPMYAOEVOSEANOTNATLAAMRVYSTSTAVMMDR1AINGVNGPG 65
DB 83 NPTFGQYLYSLRYSAEISAAVAATDSSLKAKASVANIPFTIMLDYAKV-----PD 136
OY 66 LTTYL-DAALSOQOGTTPREIVIEIVYDLPGRDCAALASNGELPATAG1QYETQYIDPI 124
DB 137 LGTYLADASSIQTKGQKQVLPIYVELPDRDCAALASNGESIDAGAEKK-DYIDQI 195
OY 125 ASILSPKYSRLRTVITIEPDSLPAVAVTNMSIQACATAVPYEGGIEYALTKLHAIPNY 184
DB 196 VPOIK--QPPDVHVAVVIEPDSLAVNLVTNLSVAKANG-GTYKASVTYALQOLSSSV-GVT 251
OY 185 IYMDAAHSGMLGMPNNAAGVQGVOKVLAASIGVNGIDGVTNTAVTTPAKPEPMATQ 244
DB 252 MYMDAGHAGMLGMPNAPIQGSEVFAMKRSADPVAFAFAFATVAYRYNALTAFA----- 305
OY 245 VGGQPVESANFYQWNPDIIDEADYAVDLYSLVAAGPSSIGMLIDTLRNGMGPPRETPG 304
DB 306 --PRPIT-----QGNPNVDEFPYIDRVPRMLKSPGPAA--FVVOGARG----- 346
OY 305 STATDVNTVFNOSKIDLRQHRGLKCNQNGAGLGQPPQASPTDFPNAHLDAYVWIKPGEIS 364
DB 347 -----QGNFRQWGMCMVNLGAGFGTGP---TSTGNDLIDALITVWRKGES 390
OY 365 DGTSAASDP-----TTGKSPKMDPYTTTSGVLTNALPMSPIAGOWFPQAFQDL 418
DB 391 DGTSSSSFRIDYSTLSVARDP-----APBAGTWFQAFYETL 431
OY 419 -ARP 421
| |
DB 432 PTRP 435
```

RESULT 12
ID 096TP4 PRELIMINARY: PRT: 465 AA.
AC 096TP4
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE CELLOBIOHYDROLASE II.
OS Pleurotus sajor-caju (Oyster mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OX Agaricales; Pleurotaceae; Pleurotus.
RN NCBI_TaxID=50053;
RP SEQUENCE FROM N.A.
RA Kearsomong S., Peberdy J.F.;
RT "Cloning a cellobiohydrolase II gene from Pleurotus sajor-caju and its
expression in Saccharomyces cerevisiae."
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY050517; M15038.1; -
DR EMBL: AY050518; M15037.1; -
KM Hydrolase.
SO SEQUENCE 465 AA; 49271 MW; 323AB1B235E265ED CRC64;

Query Match 25.1%; Score 564; DB 3; Length 465;
Best Local Similarity 31.9%; Pred. No. 2.7e-32;
Matches 151; Conservative 56; Mismatches 142; Indels 124; Gaps 16;

QY 6 NPAG-----ATFFVNPYMAOEVOSEANQTNATLAKKRVSTSTAVMMORIAI 57
DB 62 NPAGDVSHVLMCOITLSPYADVAAVAISNPALAKASVANIPIPIFEDVAKV 121
QY 58 NGVGGPGITTYDALISOOGT-TPEVIVIVYDLPGRCCALASNGELPATAGLOTY 116
DB 122 -----PTLCTYDALISTOOSTGRNOLVOIVYDLPDRCCALASNGEESIANNGIANT 175
QY 117 ETQYIDPIASISLN-----PKYSRLIVITIEPDSLPAVATNMSIOACA 160
DB 176 K-NYVQIVQIARCCPLVTSATDLACSEYQIRVAVVEPDSLANNVTMLNVEKCA 234
QY 161 TAVPYEQQIEVALTKLHAIPNYIYMDAAHSGWLGHPNNAAGVQVQVLAASGVNG 220
DB 235 GAQAAATEGVTYALQKLVN-GVYSYVDAGHAGMLGPAALGPAAQJFANLYTNAGSPSF 293
QY 221 IDGFTNTANTYPLKEPEMTATQVGQPVESANFYQWNPDIDEADY----- 267
DB 294 FRLATINAVANYNLNP-----SPDPVTS-----NANYDEIRIYNVSDCFVLWTS 340
QY 268 ---AVDLYSRVAGFPSSIGMLIDT-----LRNGMGCPNEPTGPRSTATDVNPFVQ 316
DB 341 LTICIIALAPELSSRGPRH--FIVDQGRSAVQVINGAMD----- 379
QY 317 SKIDLQNHKGLMONGAGIGQPPQASPTDFPAHLDAYVIRPGSDGTSASDPTTG 376
DB 380 -----WCNVNAGGTRP-----TSTGSSLDAIYVWRPGGSDGTS-----DTSA 421
QY 377 KKSDDPCDDPTTYSYGLTNALPNSPIA-----GQWFPQFQVLYNANPAV 423
DB 422 VRYDGHG-----LASAKRPAPAMASVSHSFOAYFEMLVANAVPAL 465

RESULT 13
ID 09CIS9 PRELIMINARY: PRT: 476 AA.
AC 09CIS9
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE AVICELASE 2.
GN AVI2.
OS Humicola insolens.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.

OX NCBI_TaxID=34413;
RN [1]
RP SEQUENCE FROM N.A.
RA Moriya T., Matanabe M., Sumida N., Okakura K., Murakami T.;
RT "Avicelase 2 from Humicola insolens: Purification, cloning, sequencing
and expression."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB048710; BAB39154.1; -
DR HSSP: P00725; IAZ6.
DR InterPro: IPR000254; CBD_fungal.
DR InterPro: IPR001524; Glyco_hydro_6.
DR Pfam: PF00734; CBD_1; 1.
DR Pfam: PF01341; Glyco_hydro_6; 1.
DR PRINTS: PR00733; GLYHDLASE6.
DR ProDom: PD001821; CBD_fungal; 1.
DR ProDom: PD003733; Glyco_hydro_6; 1.
DR SMART: SM00236; fCBD; 1.
DR PROSITE: PS00562; CBD_FUNGAL; 1.
DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
SO SEQUENCE 476 AA; 51276 MW; D2B0054F1381E653 CRC64;

Query Match 25.0%; Score 562.5; DB 3; Length 476;
Best Local Similarity 33.7%; Pred. No. 3.6e-32;
Matches 147; Conservative 54; Mismatches 146; Indels 89; Gaps 15;

QY 1 ATHTNDPYACATFFVNPYMAOEVOSEANQ-TNATLAKKRVSTSTAVMMORIAI 59
DB 113 ASTGNPFEGQIANNVYSEVHTLAIPOITDPALAAASVAEVPFQWLDNNVTVD- 171
QY 60 VNGSGPGITTYDALISO-----OQGTTPV-VIEIVIVYDLPGRCCALASNGELPATAGL 113
DB 172 -----TLVETLSRIIRANQAGANPYAAQIVYVYDLPDRCCALASNGEESIANNGIANT 223
QY 114 QTEYQYIDPIASISLNPKYSRLIVITIEPDSLPAVATNMSIOACATAVPYEQIEYA 173
DB 224 NNK-KGYINIRIELIS--FSDVRTIIVIEPDSLANNVTNNVAKCGAATYRELYIA 280
QY 174 LTKLHAIPNYIYMDAAHSGWLGHPNNAAGVQVQVLAASGVNGIDGFTNTANT- 232
DB 281 LKQILD-LPHVAMTMDACHAGMLGMPANIOPAELEPAKTYEDAGPRAVRGATVANINA 339
QY 233 -----PLKEPEMTATQVGQPVESANFYQWNPDIDEADYAVDLYSRVAGFPSSIG 285
DB 340 WSISSEPPYSP-----NPNVDEKRYIEAFRPLEARGFPAQ-- 376
QY 286 MLIDTLNMGCGPNEPTGPRSTATDVNPFVQSKIDLRQHGKMGONAGIGQPPQASPT 345
DB 377 FIVDQGRSG-----KQPTG-----OKENGCHMCNAIGTGFGRPTANT- 413
QY 346 DFPAPHLDAYVIRPGSDGTSASDPTGKKSDDPCDDPTTYSYGLTNALPNSPIAG 405
DB 414 --GHQYDAFVWVAKPGGSDGTS---DTTPARYDYHCG-----LEDALKPAPBAG 458
QY 406 QWFPAPQFQVLYNANPA 421
DB 459 QWFPAYEQVLLRNANP 474

RESULT 14
ID 09HEY8 PRELIMINARY: PRT: 471 AA.
AC 09HEY8
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE CELLOBIOHYDROLASE II.
GN CBH1 OR CBH2.
OS Trichoderma koningi.
OS Trichoderma koningii.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:22:32 ; Search time 47.58 Seconds

(without alignments)
217.151 Million cell updates/sec

Title:

US-09-917-384-4

Perfect score:

2249

1 ATHVDNYPYAGATFFVNPYWA.....ACQWEPAPDFDLVANAPAY 423

Sequence:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Scoring table:

231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters:

231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	639.5	28.4	352	US-09-286-691-26	Sequence 26, Appl
2	639.5	28.4	352	US-09-687-147-26	Sequence 26, Appl
3	555	24.7	360	US-09-286-691-25	Sequence 25, Appl
4	555	24.7	360	US-09-687-147-25	Sequence 25, Appl
5	553.5	24.6	461	US-09-118-319-9	Sequence 9, Appl
6	545	24.0	470	US-09-118-319-8	Sequence 8, Appl
7	539	24.0	457	US-09-142-759-1	Sequence 1, Appl
8	538	23.9	360	US-09-286-691-24	Sequence 24, Appl
9	538	23.9	360	US-09-687-147-24	Sequence 24, Appl
10	538	23.9	365	US-08-169-9488-12	Sequence 12, Appl
11	538	23.9	365	US-08-448-873-12	Sequence 12, Appl
12	538	23.9	365	US-08-382-452D-12	Sequence 12, Appl
13	439.5	19.5	449	US-09-118-319-7	Sequence 7, Appl
14	439.5	19.5	449	US-09-286-691-4	Sequence 4, Appl
15	439.5	19.5	449	US-09-687-147-4	Sequence 4, Appl
16	436	19.4	428	US-09-118-319-5	Sequence 5, Appl
17	434.5	19.3	326	US-09-286-691-23	Sequence 23, Appl
18	434.5	19.3	326	US-09-687-147-23	Sequence 23, Appl
19	431.5	19.2	432	US-09-118-319-2	Sequence 2, Appl
20	411	18.3	459	US-09-118-319-6	Sequence 6, Appl
21	411	18.3	459	US-09-286-691-2	Sequence 2, Appl
22	411	18.3	459	US-09-687-147-2	Sequence 2, Appl
23	242.5	10.8	291	US-09-286-691-28	Sequence 28, Appl
24	242.5	10.8	291	US-09-687-147-28	Sequence 28, Appl
25	236.5	10.5	286	US-09-286-691-27	Sequence 27, Appl
26	236.5	10.5	286	US-09-687-147-27	Sequence 27, Appl
27	231	10.3	490	US-09-109-841-2	Sequence 2, Appl

28	193.5	8.6	290	US-09-286-691-29	Sequence 29, Appl
29	193.5	8.6	290	US-09-687-147-29	Sequence 29, Appl
30	104	4.6	632	US-08-443-104-4	Sequence 4, Appl
31	104	4.6	632	US-08-442-859-4	Sequence 4, Appl
32	104	4.6	632	US-08-398-489-4	Sequence 4, Appl
33	104	4.6	632	PCT-US95-05534-4	Sequence 4, Appl
34	103	4.6	363	US-08-978-559A-2	Sequence 2, Appl
35	99.5	4.4	388	US-08-894-772-2	Sequence 2, Appl
36	99.5	4.4	388	US-09-207-844-2	Sequence 2, Appl
37	98	4.4	1042	US-08-928-361B-11	Sequence 11, Appl
38	98	4.4	1837	US-08-928-361B-5	Sequence 5, Appl
39	96.5	4.3	1751	US-09-136-574A-44	Sequence 44, Appl
40	96	4.3	363	US-09-336-601-1	Sequence 1, Appl
41	94	4.2	626	US-08-596-300A-7	Sequence 7, Appl
42	94	4.2	626	US-08-596-300A-14	Sequence 14, Appl
43	94	4.1	1004	US-09-268-347-30	Sequence 30, Appl
44	92.5	4.1	1043	US-08-928-361B-30	Sequence 30, Appl
45	92.5	4.1	1721	US-08-700-651-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-09-286-691-26
Sequence 26, Application US/09286691
Patent No. 6190189
GENERAL INFORMATION:
APPLICANT: Li, Xin-Liang
APPLICANT: Ljungdahl, Lars G.
TITLE OF INVENTION: Cellulases and Coding Sequences
FILE REFERENCE: 42-96
CURRENT APPLICATION NUMBER: US/09/286,691
CURRENT FILING DATE: 1999-04-05
EARLIER APPLICATION NUMBER: US 60/027,883
EARLIER FILING DATE: 1996-10-04
EARLIER APPLICATION NUMBER: PCT US97/18008
EARLIER FILING DATE: 1997-10-03
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 26
LENGTH: 352
TYPE: PRT
ORGANISM: Agaricus bisporus
US-09-286-691-26

Query Match 28.4%; Score 639.5; DB 4; Length 352;
Best Local Similarity 36.6%; Pred. No. 1.1e-56;
Matches 153; Conservative 62; Mismatches 134; Indels 69; Gaps 12;

QY 6 NPYAGATFFVNPYMAOYSEANOTNATLAKMRVSTYSTAVWMDRIAMINGGPG 65
||| | : : : : : ||| : : : : : ||| : : : : :
DB 4 NPYTKTWLSPFYADEVAQAADISNPSLATKASVAKIPFVFDVAVK-----PD 57
QY 66 LTTVDALDSOOGCTTPYIEIVYIDLPGRDCAALASNGELPATAGIOTYTOYIDPA 125
||| | : : : : : ||| | : : : : : ||| | : : : : :
DB 58 LGGYIADRSKNQ-----LVQIVYIDLPGRDCAALASNGELPATAGIOTYTOYIDPA 111
QY 126 SLTSPKYSRLIVYIEPDSLPNAVTNMSIOCATAPVYEGIEATLTKLHAIPNYI 185
: : : : : ||| | : : : : : ||| | : : : : : ||| | : : : : :
DB 112 AQIK--QPPDVSVAVIEPDSLANVTNINVOKCANQASAYKEGYIVAVQKINAV-GVYM 168
: : : : : ||| | : : : : : ||| | : : : : : ||| | : : : : :
QY 166 YMDAHSGLGKPNNAAGVQEVOKVNLASIVNGIDGFTVTANTYTPKEPMTATQOV 245
||| | : : : : : ||| | : : : : : ||| | : : : : :
DB 169 YIDACHAGWLGWPANISPAQQLFAQIYRDAGSPRLRGATVAVNFALRAS-----PD 220
QY 246 GGQPPESANFYQWNPDIIEADYAVDLSRLVAAAGPSSIGMLIDTLRNGWGPNEPTGS 305
: : : : : ||| | : : : : : ||| | : : : : : ||| | : : : : :
DB 221 SPDPIT-----QGSNVDIEIHTIALPRLSNAGPAH--FVDDGRSG-----262
QY 306 TATDVNTEFVNSKIDLRHGRGLMCNONGAGLQPPQASPTDPNNAHLDAYVWIKPGESD 365

```

Db 263 -----VONTRODMGDMCNKAGFGORP-----TTNGSSLIDIAIWMKPGGECD 307
Qy 366 GTSASDPTTGKSDPMCDPTTTSYGVLTNALPNSPIAGOWFPAGFDOLVANARPAY 423
Db 308 GTSNDSSP-----RFDSHCS-----LSDAHQPAPEAGTWFQAFETVLVANANPAL 352

```

RESULT 2

```

US-09-687-147-26
; Sequence 26, Application US/09687147
; Patent No. 6268198
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96a
; CURRENT APPLICATION NUMBER: US/09/687,147
; CURRENT FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US 60/027,883
; PRIOR FILING DATE: 1996-10-04
; PRIOR APPLICATION NUMBER: PCT US97/18008
; PRIOR FILING DATE: 1997-10-03
; PRIOR APPLICATION NUMBER: 09/286,691
; PRIOR FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Agaricus bisporus
US-09-687-147-26

```

```

Query Match 28.4%; Score 639.5; DB 4; Length 352;
Best Local Similarity 36.6%; Pred. No. 1,le-56;
Matches 153; Conservative 62; Mismatches 134; Indels 69; Gaps 12;

```

```

Qy 6 NPYAGATFFVNPYMAOEVOSEANOTNATLAAKRVSYSTYSAVMMDRITAINGVNGFG 65
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 4 NPYTGTVMLSPFYADEVAQAADISNPISLATYKASVAKIPFVWMDTVAKV-----PD 57
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 66 LTTYDALASOOGGTPPEVIEIYVDLPGRDCAALASNGELPATAAGLOTYETQYIDPTA 125
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 58 LGTYLDADASKNO-----LVQIVYVDLPDRCALASNGEFLANDGLNKYK-NYVDQIA 111
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 126 SILSNPKYSLRIYVITIEPDSLPAVNTNMSIQACATAVPYEAGEIEYALFKLHAIPNYI 185
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 112 AQIK--GFPDVSVAVIEPDSLANTLVNLTNOKCANASAYKEGVLYAVOKLNAV-GVIM 168
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 186 YMDAASHGWLGPNNASGYVOEVOKVLANSIGVNGIDGVTNTANTPLKEPPTATQOV 245
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 169 YIDAGAGWLGPNNASGYVOEVOKVLANSIGVNGIDGVTNTANTPLKEPPTATQOV 245
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 246 GQPYESANFYOMNPIDIDEADYAVDLSRLVAAGFPSSIGMLIDTLRNGMGPNPTGPS 305
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 221 SPDPIT-----OGNSYVDIHIETALPMLSNAGFPAH--FYDQGRSG-----262
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 306 TATDVNTFVNQSKIDLRQHRGLMCMONGAGLGOPQASPTDEPNNAHLDAVYVWIKPGEED 365
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 263 -----VONTRODMGDMCNKAGFGORP-----TTNGSSLIDIAIWMKPGGECD 307
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 366 GTSASDPTTGKSDPMCDPTTTSYGVLTNALPNSPIAGOWFPAGFDOLVANARPAY 423
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 308 GTSNDSSP-----RFDSHCS-----LSDAHQPAPEAGTWFQAFETVLVANANPAL 352
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |

```

```

RESULT 3
US-09-286-691-25
; Sequence 25, Application US/09286691
; Patent No. 6190189
; GENERAL INFORMATION:

```

```

; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96
; CURRENT APPLICATION NUMBER: US/09/286,691
; CURRENT FILING DATE: 1999-04-05
; EARLIER APPLICATION NUMBER: US 60/027,883
; EARLIER FILING DATE: 1996-10-04
; EARLIER APPLICATION NUMBER: PCT US97/18008
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Fusarium oxysporum
US-09-286-691-25

```

```

Query Match 24.7%; Score 555; DB 4; Length 360;
Best Local Similarity 33.7%; Pred. No. 4,5e-48;
Matches 142; Conservative 64; Mismatches 147; Indels 68; Gaps 14;

```

```

Qy 5 DNPYAGATFFVNPYMAOEVOSEANOTNATLAAKRVSYSTYSAVMMDRITAINGVN 61
||||| : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 3 DNPYAGVDMANNYYSEVNNLAVPRLSGAKATAAKKADVBSFO--WMDYDHSIME 59
||||| : : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 62 GGFGLTTYDALASOOGGTPPEVIEIYVDLPGRDCAALASNGELPATAAGLOTYETQYI 121
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 60 -----DTLADIRKANAG-GKYAGQFVYVDLPDRCDAALASNGEFLANDGLNKYKA-YI 112
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 122 DPLASILSNPKYSLRIYVITIEPDSLPAVNTNMSIQACATAVPYEAGEIEYALFKLHAIP 181
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 113 AKIKGILON--YSDTKVIIVIEPDSLANTLVNLTNOKCANASAYKELFYAIKELN-LP 169
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 182 NYIYMDAASHGWLGPNNASGYVOEVOKVLANSIGVNGIDGVTNTANTPLKEPPTATQ 241
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 170 NWSMYLDAGHGWLGPNNASGYVOEVOKVLANSIGVNGIDGVTNTANTPLKEPPTATQ 241
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 242 TQOVGQPYESANFYOMNPIDIDEADYAVDLSRLVAAGFPSSIGMLIDTLRNGMGPNP 301
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 221 ----GWLKSTKPDYTESNNYDEQR-INAFAPLAQBGMSNKFIVDQGRSG----KOP 271
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 302 TGPSTATDVNTFVNQSKIDLRQHRGLMCMONGAGLGOPQASPTDEPNNAHLDAVYVWIKP 361
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 272 TG-----QKAGQDMCNAKGTGFGLRPSTYTGCD--ALADAFVWVKFG 310
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 362 GTSASDPTTGKSDPMCDPTTTSYGVLTNALPNSPIAGOWFPAGFDOLVANARPAY 421
||||| : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 311 GSDGTS-----DTSARKYHCG-----LDDALKEPAPEAGTWFQAFETVLVANANP 357
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 422 A 422
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 358 S 358
||| | : : : : : | : : : : | : : : : | : : : : | : : : : |

```

```

RESULT 4
US-09-687-147-25
; Sequence 25, Application US/09687147
; Patent No. 6268198
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96a
; CURRENT APPLICATION NUMBER: US/09/687,147
; CURRENT FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US 60/027,883
; PRIOR FILING DATE: 1996-10-04
; PRIOR APPLICATION NUMBER: PCT US97/18008
; PRIOR FILING DATE: 1997-10-03

```

PRIOR APPLICATION NUMBER: 09/286,691
PRIOR FILING DATE: 1999-04-05
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 25
LENGTH: 360
TYPE: PRT
ORGANISM: Fusarium oxysporum
US-09-687-147-25

Query Match 24.7%; Score 555; DB 4; Length 360;
Best Local Similarity 33.7%; Pred. No. 4,5e-48;

Matches 142; Conservative 64; Mismatches 147; Indels 68; Gaps 14;

```
QY 5 DNPYAGATFVNDPYMAOEVOSEAA---NOTNATLAAMRVYSTSTAVWMDRIAANGVN 61
DB 3 DNPYAGVDLMANNYRSEVMNLAVPKLSGAKATAAAKAVADVSFO---WMDTYDHTSLME 59
QY 62 GGPGLTTYLDAAALSOOGGTPPEVIEIYIYDLPGRCALASNGELPATAAGLOTYETQYI 121
DB 60 -----DTLADIRKANKAG-GKYAGQFVYVDLPRRDCAAAASNGEYSLDKGANKYKA-YI 112
QY 122 DPTASILSNPKYSSLRIVTIEPDSLPAVATNNSIOACATAPVYEOGIEYALTKLHAIP 181
DB 113 AKIKGILQN--YSDTVIILVIEPDSLANTLVNLDKCAKAEASAKELYATKELN-LP 169
QY 182 NYIYMDAAHSGMLGWPNNASGYOEVOKVLNLSIGVNGIDGFTVTANTYTPLEKPFMTA 241
DB 170 NYSMTYDAGHGMGLWPANIGPAKLYAQIYKDAKPSRVRGLVTIVSNVN----- 220
QY 242 TQVGGQPVESANFYQWNPIDIDADYAVDLYSLVAAGFPSSIGMLDITLRNGGPNP 301
DB 221 ---GWLKSTKRPDYTESNPNYDQRY-INAFAPLLAQEGSNVKEFTVDGRSG-----KQP 271
QY 302 TGPSTATDVNTFVNOGSKIDLRHGLMCNONGAGLGPQASPTDPNNAHLDAYVWIKRP 361
DB 272 TG-----OKAQGDMCNKAGTGFGLRPTNTGD---ALADAFVWYKPG 310
QY 362 GESDGTSAASDPPTGKSDPMDPTTTSYGYVTNALPNSPIAGOWFPAQFDLVANARP 421
DB 311 GESDGTIS---DTSAAKYDHCG-----LDDLAKRPAEGTWQAFVQFQOLDLNANP 357
QY 422 A 422
DB 358 S 358
```

RESULT 5
US-09-118-319-9
Sequence 9, Application US/09118319
Patent No. 6114158

GENERAL INFORMATION:

APPLICANT: LI, Xin-Liang
APPLICANT: Chen, Huizhong
APPLICANT: Ljungdahl, Lars G.
TITLE OF INVENTION: Orpinomyces Cellulase Celf Protein and Coding Sequences
FILE REFERENCE: 33-98sequence listing
CURRENT APPLICATION NUMBER: US/09/118,319
CURRENT FILING DATE: 1998-07-17
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 9
LENGTH: 461
TYPE: PRT
ORGANISM: Fusarium oxysporum
US-09-118-319-9

Query Match 24.6%; Score 553.5; DB 3; Length 461;
Best Local Similarity 33.5%; Pred. No. 9,6e-48;
Matches 141; Conservative 65; Mismatches 146; Indels 69; Gaps 14;

```
QY 5 DNPYAGATFVNDPYMAOEVOSEAA---NOTNATLAAMRVYSTSTAVWMDRIAANGVN 61
DB 105 DNPYAGVDLMANNYRSEVMNLAVPKLSGAKATAAAKAVADVSFO---WMDTYDHTSLME 161
QY 62 GGPGLTTYLDAAALSOOGGTPPEVIEIYIYDLPGRCALASNGELPATAAGLOTYETQYI 121
DB 162 -----DTLADIRKANKAG-GKYAGQFVYVDLPRRDCAAAASNGEYSLDKGANKYKA-YI 214
QY 122 DPTASILSNPKYSSLRIVTIEPDSLPAVATNNSIOACATAPVYEOGIEYALTKLHAIP 181
DB 215 AKIKGILQN--YSDTVIILVIEPDSLANTLVNLDKCAKAEASAKELYATKELN-LP 271
QY 182 NYIYMDAAHSGMLGWPNNASGYOEVOKVLNLSIGVNGIDGFTVTANTYTPLEKPFMTA 241
DB 272 NYSMTYDAGHGMGLWPANIGPAKLYAQIYKDAKPSRVRGLVTIVSNVN----- 322
QY 242 TQVGGQPVESANFYQWNPIDIDADYAVDLYSLVAAGFPSSIGMLDITLRNGGPNP 301
DB 323 ---GWLKSTKRPDYTESNPNYDQRY-INAFAPLLAQEGSNVKEFTVDGRSG-----KQP 373
QY 302 TGPSTATDVNTFVNOGSKIDLRHGLMCNONGAGLGPQASPTDPNNAHLDAYVWIKRP 361
DB 374 TG-----OKAQGDMCNKAGTGFGLRPTNTGD---ALADAFVWYKPG 412
QY 362 GESDGTSAASDPPTGKSDPMDPTTTSYGYVTNALPNSPIAGOWFPAQFDLVANARP 421
DB 413 GESDGTIS---DTSAAKYDHCG-----LDDLAKRPAEGTWQAFVQFQOLDLNANP 458
QY 422 A 422
DB 459 S 459
```

RESULT 6
US-09-118-319-8
Sequence 8, Application US/09118319
Patent No. 6114158
GENERAL INFORMATION:
APPLICANT: LI, Xin-Liang
APPLICANT: Chen, Huizhong
APPLICANT: Ljungdahl, Lars G.
TITLE OF INVENTION: Orpinomyces Cellulase Celf Protein and Coding Sequences
FILE REFERENCE: 33-98sequence listing
CURRENT APPLICATION NUMBER: US/09/118,319
CURRENT FILING DATE: 1998-07-17
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 8
LENGTH: 470
TYPE: PRT
ORGANISM: Trichoderma reesei
US-09-118-319-8

Query Match 24.3%; Score 546; DB 3; Length 470;
Best Local Similarity 33.6%; Pred. No. 5,8e-47;
Matches 145; Conservative 51; Mismatches 154; Indels 82; Gaps 13;

```
QY 1 ATRVNDPYAGATFVNDPYMAOEVOSEANOTNATLAAMRVYSTSTAVWMD----- 52
DB 109 ATRVNDPYAGATFVNDPYMAOEVOSEANOTNATLAAMRVYSTSTAVWMD----- 168
QY 53 --RIAAINGVNGRGLTTYLDAAALSOOGGTPPEVIEIYIYDLPGRCALASNGELPATA 110
DB 169 EOTLADIRKANKAG--NTAG-----QPVYVDLPRRDCAAAASNGEYSLAD 212
QY 111 AGLOTYETQYIDPTASILSNPKYSSLRIVTIEPDSLPAVATNNSIOACATAPVYEOGI 170
DB 213 GGYAKYK-NYIDIRQIV--VEYSDIRTLVIEPDSLANTLVNLDKCAKAEASAKELYATKELN 269
QY 171 EYALTKLHAIPNYIYMDAAHSGMLGWPNNASGYOEVOKVLNLSIGVNGIDGFTVTANT 230
DB 270 NYAVTOLN-LPNAVYDLAGHGMGLWPANQDPAQDLFNAVYKNASSPRALRGLATNVAN 328
```

QY	231	YTPLEPMTATQOYGGOVESANFYQNNPDIDEADVNDLYSRVVAAGFPSSIGMLDT	290
Db	329	YN-----GNMTSPSTTQGNVANYENELTYHAIQPLLANNGW-SNAFTITDQ	374
QY	291	LRNGGGEPTGPSTADVNTFVNQSKIDLROHGLMCNONGAGLGOPPOASPTDFPNA	350
Db	375	GRSG--KQPLTG-----QQQWGMDCNVTIGTGFGRPSANTGD---S	409
QY	351	HLDAVWVTKRPGESDGTSAASPTTGKSKSDPCDPTTYSYCVLTNALPNSITACQWPPA	410
Db	410	LIDSFVWVKPGGECGTSDSSAP-----RFDSHC-----ALPDALOPAQAGAMFOA	456
QY	411	QEDQVLANAPRA 422	
Db	457	YFVQQLTNANPS 468	
RESULT 7			
US-09-142-759-1			
Sequence 1, Application US/09142759A			
Patent No. 6127160			
GENERAL INFORMATION:			
APPLICANT: YAMANABE, Takashi			
APPLICANT: WATANABE, Manabu			
APPLICANT: HAMAYA, Toru			
APPLICANT: SUMIDA, Naomi			
APPLICANT: AOYAGI, Kaoru			
APPLICANT: MURAKAMI, Takeshi			
TITLE OF INVENTION: PROTEIN HAVING CELLULOSE ACTIVITY AND PROCESS FOR			
FILE OF INVENTION: PRODUCING THE SAME			
FILE REFERENCE: Q51673			
CURRENT APPLICATION NUMBER: US/09/142,759A			
CURRENT FILING DATE: 1998-09-14			
EARLIER APPLICATION NUMBER: PCT/J97/00824			
EARLIER FILING DATE: 1997-03-14			
EARLIER APPLICATION NUMBER: P. HEI-8-84479			
EARLIER FILING DATE: 1996-03-14			
NUMBER OF SEQ ID NOS: 30			
SOFTWARE: Patent In Ver. 2.1			
SEQ ID NO 1			
LENGTH: 457			
TYPE: PRT			
ORGANISM: Acetomoniun cellulolyticus			
US-09-142-759-1			
Query Match			
Best Local Similarity 34.0%; Score 539; DB 3; Length 457;			
Matches 143; Conservative 56; Mismatches 157; Indels 64; Gaps 14;			
QY	6	NPYAGATFFVNVYMAQEOSEANOTNATLAKMNVSTSYSAVMMDRITAAINGVNGPG	65
Db	99	NFSGYQLYANFYSEVHTLAIPLSTGLAAATKRAETPFVWLDTRAKY-----PT	152
QY	66	LTYYL-DAAISOQGGTTPEVIEI-VIYDIPGRDCAALASNGELPATPAAGLQTYETQYIDP	123
Db	153	MCTVLANTEAARAKAGASPIAGIEFVYVYDLPDRDCAAAASNGEYFVANNQVANYKA-YIDS	211
QY	124	IASIISNPKYSLRIVTIIIEPDSLNAVYTNMSIQCATAVPYEEOGIEVALKRLAIRPV	183
Db	212	IYAOLK--AYPDVHTLIIIEPDSLNMVYTNLSSTACAEAOASAYFECVNTALINLN-LANV	268
QY	184	YIYMDAAHSGWLGMPPNNAAGVQEOVKYLNASIGVNGIDGFVYNTANTYPLKEPMTATQ	243
Db	269	AMYIDAGHAGWLGWGANLSPAQLEPATYTKKNSAPASLRLGLATINVANTY-----AM	319
QY	244	OYGGOPVESANFYQNNPDIDEADVNDLYSRVVAAGFPSSIGMLDTFLRNGGGEPTG	303
Db	320	SISPP---SYTSGDSNYDEKLYINALSPLLTSSGWMPNA-HFIMDTSHNG---VQPT-	369
QY	304	PSTADVMTFVNQSKIDLROHGLMCNONGAGLGOPPOASPTDFPNAHLDAVWVTKRPG	363

```

Db 370 -----KQAMGDMCNVIGTGFVQPTTNGD---PLEDAFVWVKPGSE 409
QY 364 SDGTSASDPPTGKSKSDPMCDPTTTSIGVLTNALPNSPIAGOWFPAQFDOLVANARPAV 423
Db 410 SDGTSNSS---AATRYDFHCG-----YSDALQAPAEAGTWPQAYFVOLLTNANPAL 456

RESULT 8
US-09-286-691-24
; Sequence 24, Application US/09286691
; Patent No. 6190189
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Li, Xingdaoh, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96
; CURRENT APPLICATION NUMBER: US/09/286,691
; CURRENT FILING DATE: 1999-04-05
; EARLIER APPLICATION NUMBER: US 60/027,883
; EARLIER FILING DATE: 1996-10-04
; EARLIER APPLICATION NUMBER: PCT US97/18008
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-286-691-24

Query Match 23.9%, Score 538, DB 4, Length 360;
Best Local Similarity 33.5%, Pred. No. 2,4e-46;
Matches 143; Conservative 50; Mismatches 152; Indels 82; Gaps 13;

QY 6 NPYGAFFGVNRYNAQEQSEANQNTNTLAKMRYSTYSAAVMD-----RIA 55
Db 4 NPEVGVTWANAAYASEVSSLAIPSLTGAMATAAAAVAKVPSEFMWLLDKTPLMEQTILA 63
QY 56 AINGVNGSGPGLTYYADALASQOQGTTPVEYIEIVYDLPGRCDAALASNGELPATAGLOT 115
Db 64 DIRANKKG--NVAQ-----QPVYVDLDEPRDCALASNGEYSIADGVAK 107
QY 116 YETQYIDDIASILSNPKYSLRIYVTLIEBDSLPANAVTKMSIOACATAVPYVEQGLEVALT 175
Db 108 YK-NYIDTRIQIV--VEYSDIRKTLIVIEBDSLANTLVNMGFPKCANAGASVIECINAVT 164
QY 176 KLHAIPNYYITMDAHSGLWLPNNAAGSYQVEQYVLNASTIVNGIDGVTYNTANTYPLK 235
Db 165 QLN-LPNNVAMYLDAAGHAGLWPAQDPAQOLFANVYKNASSPRRLRGATLVVANNY--- 220
QY 236 EPMFATQGVGQOPVESANFEYQWNPNDIDEADYADVLDYLSLVAAGFPSSIGMLIDTLRNGM 295
Db 221 -----GMNTSPSPSYTOGNAVYIEKLYIHAIGPLANHGW--SNAFFITDGRSG- 268
QY 296 GGPNEPTGPSTATDVNTEFVNOSKIDLRQHRGLMCNONGAGLGQPOASPTDEPNNAHLDAY 355
Db 269 ---KQPTG-----QQQWDMCWVIGTGFIRSANNGD---SLDSGF 304
QY 356 VWIRPGSGSDGTSASADPTTGKSKSPMCDPTTTSIGVLTNALPNSPIAGOWFPAQFDOL 415
Db 305 VWVKKGGGCGDCTSDSAP---RFDSHC-----ALPDALQAPQAGAFQAYFYOL 351
QY 416 VANARPA 422
Db 352 LTNANPS 358

RESULT 9
US-09-687-147-24
; Sequence 24, Application US/09687147
; Patent No. 6268198

```

```

; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96a
; CURRENT APPLICATION NUMBER: US/09/687,147
; CURRENT FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US 60/027,883
; PRIOR FILING DATE: 1996-10-04
; PRIOR APPLICATION NUMBER: PCT US97/18008
; PRIOR FILING DATE: 1997-10-03
; PRIOR APPLICATION NUMBER: 09/286,691
; PRIOR FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Trichoderma reesei
; US-09-687-147-24

```

```

Query Match      23.9%; Score 538; DB 4; Length 360;
Best Local Similarity 33.5%; Pred. No. 2.4e-46;
Matches 143; Conservative 50; Mismatches 152; Indels 82; Gaps 13;

```

```

QY 6 NPYAGATFFVNPYMAOEVOSEANOTNATLAKMRVSTYSTAVAMD-----RTA 55
DB 4 NPFVGTTPANNAVYSEVSLAIPSLTGAMATAAANAAYKPSMWLDTDKTPLMGTLLA 63
QY 56 AINGVNGPGLTLYIDALDSOOQGTTPVEIYIYIDLPGRCALASNGELPATAAGLOT 115
DB 64 DIRTARKNG--NYAG-----QFVYDLPDRDCAALASNGEYSTADGVAK 107
QY 116 YEQYIDPILASISNPKYSRLVITIEPDSLPNAVNTNMSIQACATAVPYEGIEYALT 175
DB 108 YK-NYIDTRIQLY--VEYSDIRTLVIEPDSLANLVTLNLTGTPCANAQSAIYECINYAVY 164
QY 176 KLHAIPNYIYIDAAHSGMLGMPNNAAGYVOEVOKVLNASIGVNGIDGFVTNTANTPLK 235
DB 165 QLN-LPNVAMYLDAGHAGMLGMPANODPPAQLFANVYKNAASPRALRGLATNVANIN--- 220
QY 236 EPPMATQOVGQPVESANFYQWNPDIADYADVLSRLVAGPSSIGMLIDTLRWG 295
DB 221 -----GNWITSPSYTQGNNAVYNEKLYIHAIIGPLANHG--SNAFFITDGRSG- 268
QY 296 GGPNEPTGSTATDVNTFYNQSKIDLROHRLGMCNONGAGLGOPPOASPTDFPNALDAY 355
DB 269 ---KQPTG-----QOQWGMCMNVIGTGFGRPSANTGD---SLDSP 304
QY 356 VWIKPGEISDGTSAASDPTTGKSDPMCDPTVYTSYGLVTNALPNSPIAGOMFPAGFDOL 415
DB 305 VWKPGEGCDGSDSSAP---RFDSHC-----ALPDALQPAQAGAMFAQATFVOL 351
QY 416 VANARPA 422
DB 352 LTNANPS 358

```

```

RESULT 10
US-08-169-948B-12
; Sequence 12, Application US/08169948B
; Patent No. 5861271
; GENERAL INFORMATION:
; APPLICANT: Fowler, Timothy
; APPLICANT: Ward, Michael
; APPLICANT: Clarkson, Kathleen
; APPLICANT: Collier, Katherine
; APPLICANT: Larens, Edmund
; TITLE OF INVENTION: No. 5861271el Cellulase Enzymes and Systems
; TITLE OF INVENTION: For Their Expression
; NUMBER OF SEQUENCES: 48

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 180 Kimball Way
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/169,948B
; FILING DATE: DEC 17 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Horn, Margaret A.
; REGISTRATION NUMBER: 33,401
; REFERENCE/DOCKET NUMBER: GC226
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7536
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 365 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-169-948B-12

```

```

Query Match      23.9%; Score 538; DB 2; Length 365;
Best Local Similarity 33.3%; Pred. No. 2.5e-46;
Matches 144; Conservative 51; Mismatches 155; Indels 82; Gaps 13;

```

```

QY 1 ATHVNPYAGATFFVNPYMAOEVOSEANOTNATLAKMRVSTYSTAVAMD----- 52
DB 4 ATSGNPFVGVTPANNAVYSEVSLAIPSLTGAMATAAANAAYKPSMWLDTDKTPLM 63
QY 53 --RIAINGVNGPGLTLYIDALDSOOQGTTPVEIYIYIDLPGRCALASNGELPATA 110
DB 64 EOTLADIRTARKNG--NYAG-----QFVYDLPDRDCAALASNGEYSTAD 107
QY 111 AGLOTETQYIDPILASISNPKYSRLVITIEPDSLPNAVNTNMSIQACATAVPYEGI 170
DB 108 GYVAKYK-NYIDTRIQLY--VEYSDIRTLVIEPDSLANLVTLNLTGTPCANAQSAIYECI 164
QY 171 EYALTKLHAIPNYIYIDAAHSGMLGMPNNAAGYVOEVOKVLNASIGVNGIDGFVTNTAN 230
DB 165 NYAVTQLN-LPNVAMYLDAGHAGMLGMPANODPPAQLFANVYKNAASPRALRGLATNVAN 223
QY 231 YTPLEKFPMTATQOVGQPVESANFYQWNPDIADYADVLSRLVAGPSSIGMLIDT 290
DB 224 YN-----GNWITSPSYTQGNNAVYNEKLYIHAIIGPLANHG--SNAFFITDQ 269
QY 291 LRNGMGPNPTGSTATDVNTFYNQSKIDLROHRLGMCNONGAGLGOPPOASPTDFPNA 350
DB 270 GRSG---KQPTG-----QOQWGMCMNVIGTGFGRPSANTGD---S 304
QY 351 HLDAYVWIKPGEISDGTSAASDPTTGKSDPMCDPTVYTSYGLVTNALPNSPIAGOMFPA 410
DB 305 LLDSEVWVKKPGEGCDGSDSSAP---RFDSHC-----ALPDALQPAQAGAMFAQ 351
QY 411 QFDQIVANARPA 422
DB 352 YFVQLLTNANPS 363

```

```

RESULT 11
US-08-448-873-12
; Sequence 12, Application US/08448873
; Patent No. 5874276

```

```

: GENERAL INFORMATION:
: APPLICANT: Fowler, Timothy
: APPLICANT: Ward, Michael
: APPLICANT: Clarkson, Kathleen
: APPLICANT: Collier, Katherine A.
: APPLICANT: Larens, Edmund
: TITLE OF INVENTION: No. 5874276el Cellulase Enzymes and Systems
: TITLE OF INVENTION: For Their Expressions
: NUMBER OF SEQUENCES: 48
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genencor International
: STREET: 180 Kimball Way
: CITY: South San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/448,873
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/169,948
: FILING DATE: 17-DEC-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Stone, Christopher L.
: REGISTRATION NUMBER: 35,696
: REFERENCE/DOCKET NUMBER: GC226D14
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 742-7555
: TELEFAX: (415) 742-7217
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 365 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-448-873-12

Query Match      23.9%; Score 538; DB 2; Length 365;
Best Local Similarity 33.3%; Pred. No. 2.5e-46;
Matches 144; Conservative 51; Mismatches 155; Indels 82; Gaps 13;

:
: 1 ATHVNDNPFYAGATFFVNPYMAQEOVSEANQTNATLAAMRVYSTYSTAVVMD----- 52
: 4 ATYSGNPFVGTTPMANAYYASEVSSLAIPSLTGAMATAAAVAKVPSMWMDLTDKPLM 63
:
: 53 --RIAINGVNGGRLTYTLDALASQOOGTPEVIEIYVDLPGRDCAALASNGELPATA 110
: 64 EQLTADITANKNG--NYAG-----QFVYIDLPRDCAALASNGEYSTAD 107
:
: 111 AGLTETFOYIDPIASISNPKYSSLRVTIIEPDSLPAVATNMSIQACATAVPYEGCI 170
: 108 GGVAKYK-NYIDITROIY--VEYSDIRTLVIEPDSLANTLVNLTGTPCANAQSAYLECI 164
:
: 171 EYALTKLHAIPNYIYMDAAHSGWMPNNAAGVQVQVYKVLNLSIGVNGIDGFTVTNAN 230
: 165 NYAVTQLN-LPNVAMTYLDAGHAGWGPANODPAAQLFANYYKNASSPRALRGLATNVAN 223
:
: 231 YTPLEKPEMTATOOVGGOPVESANFYQWNPDIIDEADVAVDLYSRLVAGFPSSIGMLIDT 290
: 224 YN-----GWMITSPPSYTQGNNAVYNEKLYIHAGPLLANGW--SNAFETIDQ 269
:
: 291 LRNGMGBNETGSTATADVTEFVNQSKIDLRQHRGLMCNONGAGLGQPPQASPTDFPNA 350
: 270 GRSG--KQPTG-----QOQMGDMCNAVIGTGFGIRPANSATD---S 304
:
: 351 HLDVYVWIKPGESEDGTSAAADPTTGKKSMDMCPPTTYSIGVLTNALPNSPIAGQWPPA 410
```

```

: 111 LLDSEVWYKPGEGCDGSDSSAP-----RFDSHC-----ALPDALQAPQAGAMPQA 351
: 411 QEDLVANARPA 422
: 352 YFVQLTNANPS 363
:
: RESULT 12
: US-08-382-452D-12
: Sequence 12, Application US/08382452D
: Patent No. 6268196
:
: GENERAL INFORMATION:
: APPLICANT: Fowler, Timothy
: APPLICANT: Clarkson, Kathleen A.
: APPLICANT: Ward, Michael
: APPLICANT: Collier, Katherine D.
: APPLICANT: Larens, Edmund A.
: TITLE OF INVENTION: NOVEL CELLULOSE ENZYMES AND SYSTEMS
: TITLE OF INVENTION: FOR THEIR EXPRESSION
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genencor International
: STREET: 925 Page Mill Road
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/382,452D
: FILING DATE: February 1, 1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Christopher L. Stone
: REGISTRATION NUMBER: 36,696
: REFERENCE/DOCKET NUMBER: GC226-2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 742-7217
: TELEFAX: (415) 742-7555
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 365 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-382-452D-12

Query Match      23.9%; Score 538; DB 4; Length 365;
Best Local Similarity 33.3%; Pred. No. 2.5e-46;
Matches 144; Conservative 51; Mismatches 155; Indels 82; Gaps 13;

:
: 1 ATHVNDNPFYAGATFFVNPYMAQEOVSEANQTNATLAAMRVYSTYSTAVVMD----- 52
: 4 ATYSGNPFVGTTPMANAYYASEVSSLAIPSLTGAMATAAAVAKVPSMWMDLTDKPLM 63
:
: 53 --RIAINGVNGGRLTYTLDALASQOOGTPEVIEIYVDLPGRDCAALASNGELPATA 110
: 64 EQLTADITANKNG--NYAG-----QFVYIDLPRDCAALASNGEYSTAD 107
:
: 111 AGLTETFOYIDPIASISNPKYSSLRVTIIEPDSLPAVATNMSIQACATAVPYEGCI 170
: 108 GGVAKYK-NYIDITROIY--VEYSDIRTLVIEPDSLANTLVNLTGTPCANAQSAYLECI 164
:
: 171 EYALTKLHAIPNYIYMDAAHSGWMPNNAAGVQVQVYKVLNLSIGVNGIDGFTVTNAN 230
: 165 NYAVTQLN-LPNVAMTYLDAGHAGWGPANODPAAQLFANYYKNASSPRALRGLATNVAN 223
:
: 231 YTPLEKPEMTATOOVGGOPVESANFYQWNPDIIDEADVAVDLYSRLVAGFPSSIGMLIDT 290
```

```

Db      224  YN-----GNIITSPSYTQGNAYVNEKLYIAIGPLLANHGW-SNAEFFIDQ 269
Oy      291  LRNGMGGPNBPTGSTATDVNFTVNOGSKIDLROHRLGCLNQNGAGLGGPPASPTDFPNA 350
Db      270  GRSG-----KQPTG-----QQQMDMCNVYGTGFGIRPSANTCD---S 304
Oy      351  HLDAYVNIKPPGESDGTSAASDPTTKKSDPMCDPTTYSYGLTNALPNSFIAGOWPFA 410
Db      305  LLDSEFVWVKKGGECDGTSDSAP---RFDSHC---ALPDALQAPAOACAMQFA 351
Oy      411  QFDOLVNAAPFA 422
Db      352  YFVOLLTNANPS 363

```

RESULT 13
US-09-118-319-7

Query Match	19.5%	Score 439.5;	DB 3;	Length 449;
Best Local Similarity	32.0%	Pred. No. 3.8e-36;		
Matches 136; Conservative	55;	Mismatches 131;	Indels 103;	Gaps 18

```

RESULT 14
US-09-286-691-4
; Sequence 4, Application US/09286691
; Patent No. 6190189
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96
; CURRENT APPLICATION NUMBER: US/09/286,691
; CURRENT FILING DATE: 1999-04-05
; EARLIER APPLICATION NUMBER: US 60/027,883
; EARLIER FILING DATE: 1996-10-04
; EARLIER APPLICATION NUMBER: PCT US97/18008
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Orpinomyces sp. PC-2
US-09-286-691-4

```

Db	366	----	RNNSAT-----	-----	WCNLKGAGIGAPQANPD--	PMPPLLDAYV	400
OY	357		WIRPGESDGTSAASDPTTGKSDPKCDPTTYSYGVLTNALPNSPIGQWPPAFQDLY				416
Db	401		WIKTPGSDSNASH-----	DPVCRN-----	SDSLQGAAPAAGSWEHDYFVLL		442
OY	417	ANAP	421				
Db	443	ENAP	447				

Query Match	19.5%;	Score 439.5;	DB 4;	Length 449;
Best Local Similarity	32.0%;	Pred. No. 3,8e-36;		
Matches 136;	Conservative	55;	Mismatches 131;	Indels 103;
			Gaps	18;

QY	2	THVDNPYGAFFVNP----	YWAQVQSEANQTNATLAAKKRVSTYSTAVAMDRIAI	57
DB	121	TTVNI:LPDSNFEFENELSYNKFQCEVDSIQRLSGSLQEKAKVKVYPTA:MLMSGAT		180
QY	58	NGVNGSPGLTITTYLDAALSOQCGTPEVEIETIYDLPGRDCAALASNGELPATAGLQTYE		117
DB	181	NEV-----ARTLNAGS-----	KTYAEVLVIMLPTKCNAGSGSNG-----GADNLSYQ	223
QY	118	TOYIDPIASILSNPKYSSLRITVITIEPDSLPAVNTNMSIOACATAVPYEEOGIEVALTKL		177
DB	224	-GYVNSIVNTIN--QYPSNRIVMIEPTDITGNLVTANNAN--CRNVHDHMKQLSVASISRF		279
QY	178	HAIPRVITYMDAAHSGRWGPRNAGGYQVQYQKVLNASIGVNGIDGFPYTNANTYPLKEP		237
DB	280	GTOKRVRYLLDAHAGWL--NSSADRAETVAETELLRNAGNCKIGISITVNSY-----		330
QY	238	FMTATQVQGVQPVESANFYQWNPDIIDEADYADVLTSRLVAAAGFPSSIGMLITLRLNGWG		297
DB	331	-----QPVYSE--YQYHQNLMRA-----LSRGV-----	RGMKRTYDTSRNG---	365
QY	298	PNEPFGSTADIVNTFFVNOSKIDLRQHRGLMCNQGAGIGQPPQASPTDFPNAH--LDAYV		356
DB	366	-----RNPSSAT-----	MCNLLGAGIGAGARPQANPD--PMPPLLDAYV	400
QY	357	WIKPGEEDGTSAASDPTTGKSDPMDCDPTTYSYGVLTNALPNSPIAGOWPFAFDOLY		416
DB	401	WIKTGESDSASSA-----	DPVCRN-----SDSLQAGAPAGSGWPHDFYVMLL	442
QY	417	ANARP 421		
DB	443	ENANP 447		

RESULT 15
US-09-687-147-4
Sequence 4, Application US/09687147
Patent No. 6268198
GENERAL INFORMATION:
APPLICANT: Li, Xin-liang
APPLICANT: Ljungdahl, Lars G.
APPLICANT: Chen, Huizhong
TITLE OF INVENTION: Cellulases and Coding Sequences
FILE REFERENCE: 42-96a

```
; CURRENT APPLICATION NUMBER: US/09/687,147
; CURRENT FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US 60/027,883
; PRIOR FILING DATE: 1996-10-04
; PRIOR APPLICATION NUMBER: PCT US97/18008
; PRIOR FILING DATE: 1997-10-03
; PRIOR APPLICATION NUMBER: 09/286,691
; PRIOR FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Orpinomyces sp. PC-2
US-09-687-147-4
```

```
Query Match 19.5%; Score 439.5; DB 4; Length 449;
Best Local Similarity 32.0%; Pred. No. 3.8e-36;
Matches 136; Conservative 55; Mismatches 131; Indels 103; Gaps 18;
```

```
QY 2 THVDNPFYAGATFEVNP---YMAQEVQSEANQTNATLAARVYSTYSTAVMMDRIAI 57
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 121 TTVNSLPTSDNEFENELYSNKKFQGEVDQSIQRLSGLSQEKAKKYYPTAAILAMSGAT 180
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 58 NGVNGGGLITTYLDAALSQOQGTPEVIEIYIYDLPGRDCAALASNGELPATAAGLOTYE 117
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 181 NEV-----ARYLNEAGS-----KTVFVFLYIMIPTRDCNAGSNG---GADNLSITYQ 223
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 118 TOYIDPIASISLNPKYSLRIVTIIIEPDSLPAATVNMSTQACATAVPYEEOGIEYALFKL 177
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 224 -GYNSTIYNTIN--QYNSRIYMIIEPDTIGNLTANNAN--CRNVHDMHKQALSTAIISKF 279
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 178 HAIPNVIYMDAHSGLWGNPNASGVYQVQKYLMSIGVNGIDGFVTNTANTYPLKEP 237
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 280 GTOKNVRVYILDAHGWML---NSSADRTAEVIAEILRNAGNGKIRGISTVSNY----- 330
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 238 FMTATQOVGGQPVESANTFYQNPIDEDYAVDLXSRLYAAGFPSSIGMLIDTLRNGWG 297
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 331 -----OPVYSE--YQYHQLNRA-----LESRGV-----RGMKEIVDTSRNG--- 365
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 298 PNEPTGPSTATDVNTFVNSKIDLRHRLMCONNGAGIGOPPOASPTDPNNAH-IDAYV 356
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 366 -----RNPSSAT-----WCNLRGAGLGRARPOANPD--FNMPLLDAYV 400
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 357 WIKPPGESDGTSAASDPTTGKSDPMCDPTYTTSYGVLTJNALPNSPIAGQWFPQAQFDQLV 416
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 401 WIKTPGESDSASAS-----DPVCRN-----SDSLQGAPAGSWFHDYFVWLL 442
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 417 ANARP 421
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 443 ENANP 447
```

Search completed: August 29, 2002, 16:22:34
Job time: 364 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:21:42 ; Search time 127 Seconds

(without alignments)
131.190 Million cell updates/sec

Title: US-09-917-384-5
Perfect score: 806
Sequence: 1 VSGIKYQYKXNDSAPGDNQ.....EQDWTKYTVYVNGVLWGTG 150

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

```

1: A.Geneseq.033802.*
2: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	370	45.9	1352	22	AAG63962
2	367	45.5	551	22	AAW18790
3	366	45.4	1350	22	AAG63963
4	365	45.3	782	12	MAR15625
5	361.5	44.9	700	12	MAR13427
6	359.5	44.6	476	21	AAV54123
7	359.5	44.6	493	20	AAV28850
8	359.5	44.6	493	20	AAV43218
9	358.5	44.5	167	17	AAV95080
10	357.5	44.4	531	18	MAR15238
11	357.5	44.4	1853	19	AAW43108

12	355	44.0	499	14	AAV42122
13	340	42.2	616	20	AAV13494
14	339	42.1	1751	20	AAV13493
15	336	41.7	1426	20	AAV13492
16	229	28.4	162	20	AAV90077
17	228	28.3	162	15	AAV63634
18	228	28.3	163	22	AAE05745
19	228	28.3	341	22	AAE05749
20	228	28.3	341	22	AAE05747
21	228	28.3	382	20	AAV39952
22	228	28.3	428	22	AAE05748
23	227	28.2	190	22	AAE05746
24	226	28.0	154	20	AAV90081
25	226	28.0	156	20	AAV90080
26	130	16.1	531	16	AAW01503
27	130	16.1	532	12	AAV13229
28	121	15.0	256	22	AAV81128
29	84.5	10.5	475	22	AAV92298
30	83	10.3	1723	22	ABG24680
31	82.5	10.2	986	21	AAV40440
32	82.5	10.2	1346	22	AAU04567
33	82.5	10.2	1346	22	AAU04581
34	82.5	10.2	1371	22	AAU04570
35	81.5	10.1	633	21	AAV84965
36	81	10.0	646	22	AAV80866
37	80	9.9	1992	17	AAW04505
38	80	9.9	1992	22	AAV69137
39	80	9.9	1992	22	AAV69133
40	80	9.9	2047	22	AAV69134
41	79.5	9.9	585	21	AAV49311
42	79.5	9.9	666	21	AAV49310
43	79.5	9.9	724	21	AAV49309
44	79	9.8	237	11	AAV05803
45	79	9.8	799	22	AAV49462

ALIGNMENTS

RESULT 1	AAV63962	standard; Protein; 1352 AA.
ID	AAV63962	
XX	AAV63962	
AC	AAV63962	
XX	29-OCF-2001 (first entry)	
DT	XX	
XX	XX	
DE	XX	Amino acid sequence of xyloglucanase enzyme.
XX	XX	xyloglucanase; family 44; glycosyl hydrolase; detergent;
KW	KW	cellulosic fiber; textile scouring.
XX	OS	Paenibacillus polymyxa.
XX	XX	
PN	XX	WO200162903-A1.
XX	PD	30-AUG-2001.
XX	PF	21-FEB-2001; 2001WO-DK00116.
XX	PR	24-FEB-2000; 2000DK-0000291.
XX	PA	(NOVO) NOVOZYMES AS.
PI	XX	Schnorr K, Jorgensen PL, Schuelein M;
XX	DR	WPI; 2001-522819/57.
XX	XX	N-PSDB; AAH75059.
PT	XX	New xyloglucanase enzyme belonging to glycosyl hydrolases family,
PT	XX	useful for detergent compositions, and textile or cellulose fiber
XX	XX	processing industries -

NK-1 cellulase. B
Truncated cellulases
Truncated cellulases
Truncated cellulases
C. cellulovorans C
Cellulose binding
Clostridium cellul
Chimeric S peptide
Clostridium cellul
Gausia luciferase
Clostridium cellul
Clostridium cellul
C. cellulovorans C
C. cellulovorans C
60 kD endoglucanase
Endoglucanase enco
C17F2 Ospa constru
C glutamicum prote
Novel human diagn
Human brain deriva
Human G-protein co
Human G-protein co
Human G-protein co
Amino acid sequenc
Bacillus lichenifo
Moraxella 200 kDa
M. catarrhalis str
M. catarrhalis str
M. catarrhalis str
Arabidopsis thalia
Arabidopsis thalia
Heat-stable endo-b
Propionibacterium

P5 Example 3; Page 81-85; 97pp; English.
XX
CC The present sequence represents a xyloglucanase of the invention. The
CC specification describes a xyloglucanase enzyme belonging to family 44
CC of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
CC of at least 30% at pH 5-8. The enzyme exhibits high performance in
CC detergent compositions and prevents binding of certain soils to the
CC xyloglucan left in the cellulosic material. It is stable at pH 5-10 at
CC room temperature and has a half life of more than 50 days when incubated
CC in a full formulated liquid detergent at 30 plus degrees celsius. The
CC enzyme is used in detergent compositions, textile industry for improving
CC the properties of cellulosic fibers, yarn, and woven or non-woven
CC fabrics, preferably in textile scouring process, and in cellulose fiber
CC processing industry for retting of fibers e.g. hemp, jute, flax, and
CC linen.
XX
SQ Sequence 1352 AA;

Query Match 45.9%; Score 370; DB 22; Length 1352;
Best Local Similarity 46.0%; Pred. No. 9,8e-28;
Matches 69; Conservative 30; Mismatches 47; Indels 4; Gaps 3;

QY 4 GLKVVYKKNDSAPGDNQIKPGLQVNTGSSVDLSTVTVRYWFTTRDGSSTLYVNCMDMA 63
DB 1200 glllyrtadtkvndhnlhpgdfqivnkgfslpineklrlyttldgdtreg-tfncdyat 1258
QY 64 MCGGNIRASFGSVNPARPTADTYIQLSFT--GGTLAAGSGTEGIQNRVKSMDMSNFTETN 121
DB 1259 lscskrngklivmekaatgadyllvslfnsdagvlapgsctgldgrlthktdwsnynead 1318
QY 122 DYSY-GTNTTFODWTKTVTYVNGVLWGMTE 150
DB 1319 dylsyk9tqlctadhpkvltlmgvltwgtc 1348

RESULT 2
AAW18790
ID AAW18790 standard; protein; 551 AA.
XX
AC AAW18790;

DT 18-NOV-1997 (first entry)
XX
DE Corrected Bacillus lautus (NCIMB 40250) endoglucanase Endo 3A.
XX
KW Endoglucanase; Endo 3A; formation; localised; variation;
KW colour density; surface; dye; fabric; family 5; cellulose;
KW hydrolysatoin; p-nitrophenyl-beta-1,4-cellobioside; stone wash;
KW blue jeans; back staining.
XX
OS Bacillus lautus.

XX WO9709410-A1.
XX
XX 13-MAR-1997.
XX
XX 03-SEP-1996; 96WO-DK00364.
XX
XX 08-SEP-1995; 95DK-0000993.
XX
XX (NOVO) NOVO-NORDISK AS.
XX
XX Flich M, Onishi M, Schulein M, Toft AH;
XX
XX WPI; 1997-192888/17.
XX

PT Localised variation of colour density in the surface of a dyed
PT cellulosic fabric - uses cellulase compsn. able to hydrolyse
PT p-nitrophenyl -beta-1,4-cellobioside
XX
PS Disclosure; Pages 15-17; 23pp; English.
XX

CC The present sequence is the corrected version of the incorrect
CC Bacillus lautus (NCIMB 40250) endoglucanase Endo 3A described in
CC WO9110732. Endo 3 can be used in novel method of forming localised
CC colour density variation on the surface of a dyed cellulosic
CC fabric. The method comprises agitating the fabric in an aqueous
CC medium (pH 6.5 to 9.0) containing a family 5 cellulase,
CC e.g. the present sequence, which can hydrolyse p-nitrophenyl-beta
CC -1,4-cellobioside, or a family 7 cellulase, and a mechanical
CC abrading agent or cellulose having abrading activity. Each
CC cellulase displays 30 % or more of its maximum activity at pH 7.
CC The process is useful to provide a stone washed look to blue jeans
CC without back staining.
XX
SQ Sequence 551 AA;

Query Match 45.5%; Score 367; DB 18; Length 551;
Best Local Similarity 47.0%; Pred. No. 6,1e-28;
Matches 71; Conservative 31; Mismatches 43; Indels 6; Gaps 4;

QY 2 SSGLKVVYKKNDSAPGDNQIKPGLQVNTGSSVDLSTVTVRYWFTTRDGSSTLYVNCMD 61
DB 402 tgnlvvykvgtlsatdnqmkpsfnlknngtltpvnlsglklrlyftkd-gradmasafdw 460
QY 62 AAMGCGNIRASFGSVNPARPTADTYIQLSFT--GGTLAAGSGTEGIQNRVKSMDMSNFT 119
DB 461 aqiganvsaat--anfigsntdlyvelsfasgsislpa99tqldqlrmyktdwsnlne 518
QY 120 TNDYSY-GTNTTFODWTKTVTYVNGVLWGMGT 149
DB 519 andysydgaktayadwrvltlmgngtlwgt 549

RESULT 3
AAG63963
ID AAG63963 standard; protein; 1350 AA.
XX
XX AAG63963;

DT 29-OCT-2001 (first entry)
XX
DE Amino acid sequence of xyloglucanase enzyme.
XX

KW xyloglucanase; family 44; glycosyl hydrolase; detergent;
KW cellulosic fiber; textile scouring.
XX

OS Paenibacillus polymyxa.

XX
XX Key location/Qualifiers
XX Misc-difference 1347 /note= "Glu encoded by GAACG"
XX

XX WO200162903-A1.
XX

XX 30-AUG-2001.
XX

XX 21-FEB-2001; 2001WO-DK00116.
XX

XX 24-FEB-2000; 2000DK-0000291.
XX

XX (NOVO) NOVOZYMES AS.
XX

XX Schnorr K, Jorgensen PL, Schulein M;
XX

XX WPI; 2001-522819/57.
XX

XX N-PSDB; AAH75060.
XX

PT New xyloglucanase enzyme belonging to glycosyl hydrolases family,
PT useful for detergent compositions, and textile or cellulose fiber
PT processing industries -
XX
PS Disclosure; Page 86-89; 97pp; English.
XX

PR 06-MAY-1998; 98US-0073684.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Andersen LN, Bjornvad ME, Lange NEK, Schnorr K, Schuelein M;
 XX WPI; 1999-610578/52.
 DR N-PSDB; AAX90978.
 XX
 PT New isolated pectate lyase enzymes -
 XX
 PS Example 4; Page 85-86; 93pp; English.
 XX
 CC The present sequence is a pectate lyase-linker-CBD fusion protein.
 CC Plasmid pMB914 was constructed using pectate lyase gene from
 CC *Bacillus licheniformis*, ATCC 14580 and cellulose binding domain (CBD)
 CC gene from *Clostridium thermocellum* YS fused with a linker molecule.
 CC *Bacillus subtilis* was transformed with plasmid pMB914 for expression of
 CC the fusion protein. Pectate lyase can be used in detergent compositions,
 CC for cleaning hard surfaces, for machine treatment of fabrics, for
 CC improving the properties of cellulosic fibres, yarn, woven or non-woven
 CC fabric, for the degradation of plant material e.g. recycled waste paper,
 CC mechanical paper-making pulps or fibres subjected to retting process, for
 CC preparing animal feed and for processing wine or juice.
 CC
 XX Sequence 493 AA;
 SO
 Query Match 44.6%; Score 359.5; DB 20; Length 493;
 Best Local Similarity 44.0%; Pred. No. 2.9e-27;
 Matches 70; Conservative 34; Mismatches 44; Indels 11; Gaps 3;
 OY 1 VSGGLKQVKNNDAPDNOIKPGLQVNTGSSVSDSTVRYWTRDGGSSSTLVYVNC 60
 DB 331 vsqnlkvefynspdstltnpqlfvtntgsaidsklitlyryevdgqkqdtf-cd 389
 OY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQLSFTGGLAAGSGTGEIONRVNK 111
 DB 390 haallgngsyngltanvkgtfvmssstnnadlyleisftggtlpgahv-qiggrfak 448
 OY 112 SDMSNFTENDSYGNTTFODWTKRYVYVNGVLMWCTE 150
 DB 449 ndwsnytgndysfksrsgfvewdqvcaylmgvlywgke 487
 RESULT 8
 AAY43218
 ID AAY43218 standard; Protein; 493 AA.
 XX
 AC AAY43218;
 XX
 DT 13-JAN-2000 (first entry)
 XX
 DE Pectate lyase CBD fusion protein sequence.
 XX
 KW Pectate lyase; polysaccharide lyase; enzyme; pectin degradation;
 KW polygalacturonide; detergent composition; hard surface treatment;
 KW cellulosic fibre; plant material degradation; recycled waste paper;
 KW mechanical paper-making pulp; wine processing; cellulose binding domain;
 KW ClpB.
 XX
 OS *Clostridium thermocellum*.
 OS *Bacillus* sp.
 OS Synthetic.
 XX
 PN W09927084-A1.
 XX
 PD 03-JUN-1999.
 XX
 PF 24-NOV-1998; 98NO-DK00515.
 XX
 PR 24-NOV-1997; 97DK-0001343.
 PR 24-NOV-1997; 97DK-0001344.

PR 06-MAY-1998; 98US-0073684.
 PR 02-NOV-1998; 98US-0184217.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Andersen LN, Schuelein M, Lange NEK, Bjornvad ME, Moller S;
 PI Glad SOS, Kauppinen MS, Schnorr K, Kongsbaek L;
 XX WPI; 1999-610579/52.
 DR N-PSDB; AA231562.
 XX
 PT New isolated pectate lyase enzymes -
 XX
 PS Claim 26; Page 106-108; 113pp; English.
 XX
 CC This sequence is a fusion protein comprising a *Bacillus* species
 CC pectate lyase of the invention, fused via a linker to the *Clostridium*
 CC thermocellum ClpB protein cellulose binding domain (CBD). The pectate
 CC lyase enzymes are obtained from *Bacillus agardhaerens* (BA), *Bacillus*
 CC *licheniformis* (BL), *Bacillus halodurans* (BH) and other *Bacillus* species.
 CC The pectate lyase enzymes can be used for degrading pectin, pectate and
 CC polygalacturonides. They can be used in detergent compositions, for
 CC cleaning hard surfaces, for machine treatment of fabrics, for improving
 CC the properties of cellulosic fibres, yarn, woven or non-woven fabric, for
 CC the degradation of plant material (e.g. recycled waste paper, mechanical
 CC paper-making pulps or fibres subjected to a retting process, for
 CC preparing animal feed or for processing wine or juice. DNA encoding the
 CC enzymes can also be used for the production of transgenic plants.
 CC
 XX Sequence 493 AA;
 SO
 Query Match 44.6%; Score 359.5; DB 20; Length 493;
 Best Local Similarity 44.0%; Pred. No. 2.9e-27;
 Matches 70; Conservative 34; Mismatches 44; Indels 11; Gaps 3;
 OY 1 VSGGLKQVKNNDAPDNOIKPGLQVNTGSSVSDSTVRYWTRDGGSSSTLVYVNC 60
 DB 331 vsqnlkvefynspdstltnpqlfvtntgsaidsklitlyryevdgqkqdtf-cd 389
 OY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQLSFTGGLAAGSGTGEIONRVNK 111
 DB 390 haallgngsyngltanvkgtfvmssstnnadlyleisftggtlpgahv-qiggrfak 448
 OY 112 SDMSNFTENDSYGNTTFODWTKRYVYVNGVLMWCTE 150
 DB 449 ndwsnytgndysfksrsgfvewdqvcaylmgvlywgke 487
 RESULT 9
 AAR95080
 ID AAR95080 standard; peptide; 167 AA.
 XX
 AC AAR95080;
 XX
 DT 30-OCT-1996 (first entry)
 XX
 DE Cellulose binding domain.
 XX
 KW Cellulose binding domain; CBD; hapten; moiety; biotin; avidin;
 KW streptavidin; affinity chromatography; cell separation;
 KW cell immobilisation; protein immobilisation; enzyme immobilisation;
 KW multi-enzyme reactors; signal immunoassays; drug delivery; pesticide;
 KW cellulose; chitin.
 XX
 OS *Clostridium thermocellum*.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 139 /note="Unidentified amino acid."
 XX
 PN W09613524-A1.
 XX

```

PD 09-MAY-1996.
XX
XX 26-OCT-1995; 95WO-US13813.
XX
XX 27-OCT-1994; 94IL-0111415.
XX
XX (TECR ) TECHNION RES & DEV FOUND LTD.
XX (UYRA-) UNITA RAMOT APPLIED RES & IND DEV LTD.
XX (YEDA ) YEDA RES & DEV CO LTD.
XX
XX Bayer EA, Lamed R, Morag E, Shoham Y, Wilchek M;
XX
XX WPI; 1996-239453/24.
XX
XX Modified cellulose-binding domain mols. - having attached hapten,
XX PT partic, biotin, to provide an affinity system for, e.g. seps.,
XX PT assays, reactors, delivery etc.
XX
XX Claim 12; Page 34-35; 53pp: English.
XX
XX A modified cellulose binding domain (CBD) or fraction of it, linked
XX CC to a hapten moiety via one or more Cys or Lys residues and complexes
XX CC comprising the biotinylated CBD and a biotin-binding molecule
XX CC selected from modified or unmodified avidin or streptavidin or an
XX CC anti-biotin antibody, can be used in affinity chromatography, cell
XX CC separation, cell, protein and enzyme immobilisation, selective
XX CC removal of biological materials, multienzyme reactors, signal
XX CC immunoassays and drug (e.g. pesticide) delivery. The CBD molecule
XX CC can be modified with the hapten without affecting its high affinity
XX CC for cellulose, or chitin.
XX
XX Sequence 167 AA;
XX
XX
XX Query Match 44.5%; Score 358.5; DB 17; Length 167;
XX Best Local Similarity 44.0%; Pred. No. 9.1e-28;
XX Matches 70; Conservative 34; Mismatches 44; Indels 11; Gaps 3;
XX
XX 1 VSGGLKVOYKKNDSAPGDNQIKRGLOLVNTGSSVDLSYTVRYWFTTRGGSSSTLYVNCND 60
XX ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
XX 5 vsnqlkvefynspdtlnspqfivntgssaldsklrlrytyvdygqkqdfw-cd 63
XX
XX 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQLSFTGTTLAAGSGTEIGIONRVNK 111
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
XX 64 haaligsngsyngitsnkvgtfvmssstnnadtyleisfgtglepgahv-qiggrfak 122
XX
XX 112 SDMSNFTETNDYSYGTNTFTFDQMTKYTVYVNGVLWGTE 150
XX : ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 123 ndwsnytgndysfkssqfvewdqvtaylngvltwvke 161
XX
XX
XX RESULT 10
XX AAW15238
XX ID AAW15238 standard; protein; 531 AA.
XX
XX AAW15238;
XX
XX 28-JAN-1998 (first entry)
XX
XX Scaffoldin protein from Clostridium thermocellum.
XX
XX Docketin; CelD; CelS; CIPa; scaffoldin; cellulose binding domain;
XX KW Chromatographic separation; soluble substrate modification; CBD;
XX KW multi-enzyme delivery system; animal feed; paper production;
XX KW plant protection; pest control.
XX
XX Clostridium thermocellum.
XX
XX Key Location/Qualifiers
XX FH 1.153
XX FT Region /label= internal_repeat_element_1
XX FT 154...306
XX FT Region /label= internal_repeat_element_2

```

```

FT Domain 239..531
FT /label= cellulose-binding
FT /note= "only 60% of the CBD is present"
XX
XX WO9714789-A2.
XX
XX 24-APR-1997.
XX
XX 16-OCT-1996; 96WO-US16485.
XX
XX 17-NOV-1995; 95US-0559968.
XX PR 17-OCT-1995; 95US-0005701.
XX
XX (GENV ) GENENCOR INT INC.
XX
XX Bolt RR, Clarkson KA, Fowler T, Iiu C, Ward M, Xia H;
XX
XX WPI; 1997-245106/22.
XX
XX Composition with enzymes non-covalently bound to a peptide backbone
XX PT - used as a multi-enzyme delivery system, e.g. in food processing,
XX PT textiles and pest control
XX
XX Claim 10; Fig 6; 29pp: English.
XX
XX This protein sequence represents a scaffoldin based on the CIPa protein
XX CC of Clostridium thermocellum. The scaffoldin protein is used in a novel
XX CC composition that comprises at least 2 enzymes non-covalently bound to a
XX CC peptide backbone (i.e. present sequence). The scaffoldin comprises a
XX CC number of internal repeating units and at least one cellulose binding
XX CC domain (CBD). The CBD may be altered to modify its affinity for
XX CC cellulose, which may be desirable where cellulose binding would be
XX CC disadvantageous. The enzyme is bound to a repeating element of the
XX CC region of the enzyme, which binds to a repeating element of the
XX CC scaffoldin. The dockerin is preferably CelS (AAW15237) or CelD
XX CC (AAW15236). The composition can be used in reducing allergenicity,
XX CC producing synergistic effects, and facilitating selective modification
XX CC of substrate. By taking advantage of the cellulose binding domain of the
XX CC complex, the complex could be immobilised for use in chromatographic
XX CC separations or for soluble substrate modification. By adding the
XX CC scaffoldin domain, it is possible to recover enzymes, or to quantify
XX CC the amount of an enzyme in a solution. The composition could also be
XX CC used in a multi-enzyme delivery system which could be used in the food
XX CC industry, in food processing, animal feed, textiles, bioconversion,
XX CC pulp and paper production, plant protection and pest control, as a wood
XX CC preservative, topical lotions, and biomass conversions.
XX
XX Sequence 531 AA;
XX
XX
XX Query Match 44.4%; Score 357.5; DB 18; Length 531;
XX Best Local Similarity 44.0%; Pred. No. 5.1e-27;
XX Matches 70; Conservative 34; Mismatches 44; Indels 11; Gaps 3;
XX
XX 1 VSGGLKVOYKKNDSAPGDNQIKRGLOLVNTGSSVDLSYTVRYWFTTRGGSSSTLYVNCND 60
XX ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
XX 312 vsnqlkvefynspdtlnspqfivntgssaldsklrlrytyvdygqkqdfw-cd 370
XX
XX 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQLSFTGTTLAAGSGTEIGIONRVNK 111
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
XX 371 haaligsngsyngitsnkvgtfvmssstnnadtyleisfgtglepgahv-qiggrfak 429
XX
XX 112 SDMSNFTETNDYSYGTNTFTFDQMTKYTVYVNGVLWGTE 150
XX : ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 430 ndwsnytgndysfkssqfvewdqvtaylngvltwvke 468
XX
XX
XX RESULT 11
XX AAW43108
XX ID AAW43108 standard; protein; 1853 AA.
XX
XX AAW43108;
XX

```



```
XX OS Unidentified.
XX XX
XX PN EP921188-A2.
XX PD 09-JUN-1999.
XX PF 15-SEP-1998; 98EP-0810919.
XX PR 19-SEP-1997; 97US-0932571.
XX PA (CLRN ) CIARIANT FINANCE BVI LTD.
XX PI Anderson P, Bergquist PL, Daniels RM, Farrington GK;
XX PI GIBBS MD, Morgan H, Williams DP;
XX DR WPI: 1999-315403/27.
XX DR N-PSDB: AAX55660.
XX PT New truncated cellulase proteins, useful in detergents and for
XX PT producing 'stonewashed' denim
XX PS Claim 7; Page 42-43; 65pp; English.
XX PS
XX CC The invention relates to a recombinant cellulase active protein free of
XX CC proteases of native thermophilic and alkaliphilic origin, comprising
XX CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
XX CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
XX CC length sequences, or functional equivalents. Cel B5 extends from amino
XX CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino
XX CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends
XX CC from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel
XX CC E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to
XX CC K1751 and the stability region extends from amino acid E482 to G635 in
XX CC the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
XX CC enzymes are useful in laundry detergent compositions to prevent or
XX CC enzymatic staining, backstaining or graying, for use on cellulosic
XX CC materials including cotton-containing fabrics. They are especially useful
XX CC for preventing redeposition of colorant during stonewashing, and for
XX CC processing of textiles where cellulose breakdown is required. The new
XX CC truncated enzymes show reduced redeposition of dye compared to using
XX CC non-truncated cellulase compositions.
XX CC
XX CC Sequence 616 AA;
XX SQ
XX
XX Query Match 42.2%; Score 340; DB 20; Length 616;
XX Best Local Similarity 43.8%; Pred. No. 3.4e-25;
XX Matches 67; Conservative 30; Mismatches 52; Indels 4; Gaps 3;
XX
XX QY 1 VSGGKQVYKNNDSAPGDNQIKPGLQVNTGSSVDSLTVYRYKFTRDGSSSTLVYKND 60
XX :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 1 mgsykvlykmetasqstirpwtfkivngssvdlstrvklrlytvdgddpqsav-cd 59
XX
XX QY 61 WAAMCGNIRASFGSVNPATP7ADTYLQLSFTGCT--LAAGSGTEIGIONRVKSDMSNPT 118
XX :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 60 waqiaasvltfivkissgvsqadyilevgfssgagqlqpgkdtgdiqyrfnkndwsyn 119
XX
XX QY 119 ETNDYSY-GTNTTFQDMTKVTVYVNGVLWGTGTE 150
XX :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 120 qaddswslqsmtnlygenakvltlyvgvlywqge 152
XX
XX RESULT 14
XX ID AAY13493
XX AC AAY13493;
XX DT 30-JUL-1999 (first entry)
XX DE Truncated cellulases comprising amino acid sequence.
XX XX
```

```
KW KM Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
KW KM Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
KW KM cotton-containing fabric; stonewashing.
XX OS Unidentified.
XX XX
XX PN EP921188-A2.
XX PD 09-JUN-1999.
XX PF 15-SEP-1998; 98EP-0810919.
XX PR 19-SEP-1997; 97US-0932571.
XX PA (CLRN ) CIARIANT FINANCE BVI LTD.
XX PI Anderson P, Bergquist PL, Daniels RM, Farrington GK;
XX PI GIBBS MD, Morgan H, Williams DP;
XX DR WPI: 1999-315403/27.
XX DR N-PSDB: AAX55662.
XX PT New truncated cellulase proteins, useful in detergents and for
XX PT producing 'stonewashed' denim
XX PS Claim 7; Page 37-41; 65pp; English.
XX PS
XX CC The invention relates to a recombinant cellulase active protein free of
XX CC proteases of native thermophilic and alkaliphilic origin, comprising
XX CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
XX CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
XX CC length sequences, or functional equivalents. Cel B5 extends from amino
XX CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino
XX CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends
XX CC from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel
XX CC E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to
XX CC K1751 and the stability region extends from amino acid E482 to G635 in
XX CC the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
XX CC enzymes are useful in laundry detergent compositions to prevent or
XX CC enzymatic staining, backstaining or graying, for use on cellulosic
XX CC materials including cotton-containing fabrics. They are especially useful
XX CC for preventing redeposition of colorant during stonewashing, and for
XX CC processing of textiles where cellulose breakdown is required. The new
XX CC truncated enzymes show reduced redeposition of dye compared to using
XX CC non-truncated cellulase compositions.
XX CC
XX CC Sequence 1751 AA;
XX SQ
XX
XX Query Match 42.1%; Score 339; DB 20; Length 1751;
XX Best Local Similarity 44.7%; Pred. No. 1.6e-24;
XX Matches 67; Conservative 29; Mismatches 50; Indels 4; Gaps 3;
XX
XX QY 4 GLKVOYKNNDSAPGDNQIKPGLQVNTGSSVDSLTVYRYKFTRDGSSSTLVYKNDMA 63
XX :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 678 gvkvyklykmetasqstirpwtfkivngssvdlstrvklrlytvdgddpqsav-cdwaq 736
XX
XX QY 64 MCGNIRASFGSVNPATP7ADTYLQLSFTGCT--LAAGSGTEIGIONRVKSDMSNPTETN 121
XX :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 737 lqasvltfivkissgvsqadyilevgfssgagqlqpgkdtgdiqyrfnkndwsynqad 796
XX
XX QY 122 DYSY-GTNTTFQDMTKVTVYVNGVLWGTGTE 150
XX :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 797 dswslqsmtnlygenakvltlyvgvlywqge 826
XX
XX RESULT 15
XX ID AAY13492
XX AC AAY13492;
XX DT 30-JUL-1999 (first entry)
XX DE
```


This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:24:03 ; Search time 72.78 Seconds
(without alignments)
198,041 Million cell updates/sec

Title: US-09-917-384-5
Perfect score: 806
Sequence: 1 VSGIKVQYKNNDSAPGDNQ.....FQDWTKVTYVYNGVLWGT 150

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	380	47.1	145	2	cellulase homolog
2	364	45.2	508	2	cellulase (EC 3.2.
3	362	44.9	499	2	cellulase (EC 3.2.
4	361.5	44.9	700	2	cellulase (EC 3.2.
5	357.5	44.4	1854	2	cellulase (EC 3.2.
6	357	44.3	508	2	cellulase (EC 3.2.
7	355	44.0	499	2	cellulase (EC 3.2.
8	348.5	43.2	1711	2	cellulase (EC 3.2.
9	341	42.3	1719	2	cellulase (EC 3.2.
10	336	41.7	1331	2	cellulase (EC 3.2.
11	335.5	41.6	504	2	cellulase (EC 3.2.
12	335	41.6	915	2	cellulase (EC 3.2.
13	335	41.6	1039	2	cellulase (EC 3.2.
14	335	41.6	1742	2	cellulase (EC 3.2.
15	335	40.3	505	2	cellulase (EC 3.2.
16	315	39.1	486	2	cellulase (EC 3.2.
17	286	35.5	986	2	cellulase (EC 3.2.
18	282.5	35.0	1483	2	cellulase (EC 3.2.
19	270	33.5	1162	2	cellulase (EC 3.2.
20	265	32.9	586	2	cellulase (EC 3.2.
21	254.5	31.6	1230	2	cellulase (EC 3.2.
22	253	31.4	879	2	cellulase (EC 3.2.
23	228	28.3	1848	2	cellulase (EC 3.2.
24	112	13.9	1045	2	cellulase (EC 3.2.
25	94.5	11.7	1428	2	cellulase (EC 3.2.
26	92.5	11.5	269	2	cellulase (EC 3.2.
27	90.5	11.2	698	2	cellulase (EC 3.2.
28	90.5	11.2	698	2	cellulase (EC 3.2.
29	90.5	11.2	725	2	cellulase (EC 3.2.

30	89.5	11.1	636	2	cellulase (EC 3.2.
31	88.5	11.0	698	2	cellulase (EC 3.2.
32	87.5	10.9	319	2	cellulase (EC 3.2.
33	87.5	10.9	712	2	cellulase (EC 3.2.
34	87.5	10.9	2348	2	cellulase (EC 3.2.
35	85.5	10.6	1910	2	cellulase (EC 3.2.
36	85	10.5	268	2	cellulase (EC 3.2.
37	85	10.5	582	2	cellulase (EC 3.2.
38	84.5	10.5	5188	2	cellulase (EC 3.2.
39	84.5	10.5	5291	2	cellulase (EC 3.2.
40	84	10.4	781	2	cellulase (EC 3.2.
41	82.5	10.2	618	2	cellulase (EC 3.2.
42	82	10.2	1649	2	cellulase (EC 3.2.
43	81.5	10.1	491	2	cellulase (EC 3.2.
44	81.5	10.1	2204	2	cellulase (EC 3.2.
45	81	10.0	238	1	cellulase (EC 3.2.

ALIGNMENTS

RESULT 1

A41897
cellulase homolog - Bacillus lautus (fragment)
C/Species: Bacillus lautus
C/Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 15-Oct-1999
C/Accession: A41897; S27498
R/Hansen, C. K.; Diderichsen, B.; Jorgensen, P. L.
J. Bacteriol. 174, 3522-3531, 1992
A/Title: cels from Bacillus lautus P1236 encodes a novel cellulose-binding endo-beta-
A/Reference number: A41897; MUID:92276330
A/Accession: A41897
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-145 <HAN>
A/Cross-references: EMBL:M76586; NID:g142661; PID:AAA22302.1; PID:g142662
A/Experimental source: PL236
A/Note: sequence extracted from NCBI backbone (NCBIP:104604)

Query Match	Score	DB 2:	Length	145:
Best Local Similarity	50.06	Pred. No. 3.1e-27		
Matches 73; Conservative 29; Mismatches 38; Indels 6; Gaps 4;				
OY 7 VOYKNDSPGDNQIKFGLVNTGSSVDSTVRYMFRDGGSTLYVNCDAWAGC 66				
DB 1 LQYAAATNADNDQIKSFNKNNGTSNVDSTIKIRYFKDSAAVNGH-IDMAQLG 59				
OY 67 GNIRASFGSVNPAFPTADYTLQLSFT--GGTLAAGSTGELQNRVNSDMSNFTETNDYS 124				
DB 60 SNIGISFG--NHTGNSDTLYVELSFSSSEAGSIAGGSGGTQLMNSKTDMSNFTETNDYS 117				
OY 125 Y-GTNTTFQDWTKVTYVYNGVLWGT 149				
DB 118 FDGTKFAFADMDRVLYVYNGOIVWGT 143				

RESULT 2

G69593
cellulase (EC 3.2.1.4) bgIC precursor - Bacillus subtilis
N/Alternate names: endo-1,4-beta-glucanase
C/Species: Bacillus subtilis
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000
C/Accession: G69593; A26114; I40353; S24239; S49103; I39803
R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emerson, P.T.; Ertlan, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holzapfel, S.; Hosono, S.; Hullo, M
koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivoletta, C.; Rocha, E.; Rochie B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
A. Authors: Schlecht, S.; Schroeter, R.; Scroffone, A.; Sekiguchi, J.; Sekowska, A.; Serd
kechul, M.; Tamakoshi, A.; Tanaka, T.; Terstra, P.; Toniguchi, A.; Tostato, V.; Uchiyama,
T.; Winters, P.; Wipst, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A. Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Zanchin, A.
A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A:Reference number: A69580; MUID:98044033
A:Accession: 669593
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-508 <KUN>
A:Cross-references: GB:J99113; GB:AL009126; NID:g2634090; PIDN:CA813696.1; PID:e1183471.
A:Experimental source: Strain 168
R:McKay, R.M.; Lo, A.; Willick, G.; Zuker, M.; Baird, S.; Dove, M.; Moraneli, F.; Sel
Nucleic Acids Res. 14, 9159-9170, 1986
A>Title: Structure of a *Bacillus subtilis* endo-beta-1,4-glucanase gene.
A:Reference number: A26114; MUID:87066783
A:Accession: A26114
A:Molecule type: DNA
A:Residues: 10-508 <MAC>
A:Cross-references: GB:X04689; NID:g39823; PIDN:CAA28392.1; PID:g39824
A:Experimental source: Strain PAP115
A>Note: part of this sequence, including the amino end of the mature form, was confirmed
R:Lindahl, V.; Aa, K.; Tronamo, A.
A>Title: Nucleotide sequence of an endo-beta-1,4-glucanase gene from *Bacillus subtilis* C
A:Reference number: I40353; MUID:95225656
A:Accession: I40353
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 10-291, 'N', 293-508 <LIN1>
A:Cross-references: EMBL:X67044; NID:g39776; PIDN:CAA7429.1; PID:g39777
R:Lindahl, V.; Aa, K.
submitted to the EMBL Data Library, June 1992
A:Reference number: S24239
A:Accession: S24239
A:Molecule type: DNA
A:Residues: 10-291, 'N', 293-508 <LIN2>
A:Cross-references: EMBL:X67044; NID:g39776; PIDN:CAA7429.1; PID:g39777
A:Experimental source: strain CK-2
R:Wolf, M.; Geczli, A.; Borstis, R.
submitted to the EMBL Data Library, December 1993
A:Description: Genes encoding beta glucan-hydrolyzing enzymes in *Bacillus subtilis*: cons
A:Reference number: S49103
A:Accession: S49103
A:Molecule type: DNA
A:Residues: 10-508 <WOL>
A:Cross-references: EMBL:Z29076; NID:g509266; PIDN:CAA82317.1; PID:g509267
R:Seo, Y.S.; Lee, Y.H.; Pek, U.H.; Kang, H.
Korean J. Microbiol. 24, 235-242, 1986
A>Title: Analysis on the nucleotide sequence of the signal region of *Bacillus subtilis*
A:Reference number: I39803
A:Accession: I39803
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 10-13, 'V', 15, 'I', 17-18, 'V', 20-21, 'F', 23, 'V', 25-26, 'AT', 29-31, 'P', 33, 'PQ', 36
A:Cross-references: GB:M8634; NID:g142657; PIDN:AAA22300.1; PID:g142658
A:Experimental source: Strain ATCC 6633
C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal prote
C:Genetics:
A:Gene: bglC
A:Function:
A:Pathway: cellulose degradation
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
I:38/Domains: signal sequence #status predicted <IG>
I:39-508/Product: cellulase #status predicted <AT>

Query Match	45.2%;	Score 364;	DB 2;	Length 508;
Best Local Similarity	44.6%;	Pred. No. 3.7e-25;		
Matches 66;	Conservative 32;	Mismatches 46;	Indels 4;	Gaps 2

OY 4 GLKQVKNNDSPGNDAPNKKPELOLVNTGGSSVDLSTJVAWRF--TPOGSGTLVYNCDMA 62
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 362 GISIVQRYAGSDSAMSNOIRPOLQIKNNNGTTVDLKDVTAARYAKKNGQN---FDGDYA 418

OY 63 AMGCNNIRASFGSVNPATPTADTYLOLSFTGGFLAAGSGTCGEIONRVNKSDMSNFETEND 122
||||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 419 QIGCGNVTHKEVTLLHKPKOGADFYLELGTFKKNGTLPAGASTGNIOLRILHDDMSNYAOSGD 478

Oy 123 YSYGNTFFODMTKYVVYVNGVLWGTE 150
|| : || : || : || : || : || : || : || : || : || : || : || : || : || : || : || :
Db 479 YSFERSNTEFKTKITLTLYDGSKLIWGTE 506

RESULT 3
JN0111 cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain BSE616)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C>Date: 17-Apr-1993 #sequence-revision 17-Apr-1993 #text-change 28-May-1999
C:Accession: JN0111
R:Park, S.H.; Kim, H.K.; Pack, M.Y.
A:Title: Characterization and structure of the cellulase gene of Bacillus subtilis BS
A:Reference number: JN0111; MUID:91299280
A:Accession: JN0111
A:Molecule type: DNA
A:Residues: 1-499 <PAR>
A:Cross-references: GB:O01057; NID:g216387; PIDN:BAA00859.1; PID:d1001323; PID:g216388
C>Note: The authors translated the codon ATA for residue 102 as Tyr
C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal
C:Function:
A>Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
C:KeyWords: domain; signal sequence #status predicted <SIG>

```

Query Match      44.9%  Score 362;  DB 2;  Length 499;
Best Local Similarity 44.2%  Pred. No. 5.5e-25;
Matches 65;  Conservative 29;  Mismatches 51;  Indels 2;  Gaps 1;

OY      4  GLKVOYKKNNDAPGDNQIKPGLQLVNTGSSVDLSSTVVRKWFTRDGGSTPLVYNCWAA 63
      |||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 353  GISVYRAGDSGMSNSNOGRPOLQIKNNNTVYDLKDVYARKWYAKNKNQNV--DCDVAQ 410

OY      64  MCGGNIRASFGSVNPAFTADTVYQLSFTGCTLAAGSGTGEIQNRVKSQMSNFTETNDY 123
      |||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 411  LGCGNVYKFFKTLHKPKOGADTVYELGFKNGTLLAPGASTGNIQRLHNDWMSNQAQSGDY 470

OY      124  SYGTNTFQDWTKTVVYVNGVLWNGTE 150
      |:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 471  SFKSNFTKTKTKITLIVDQKLIMGTE 497

RESULT      4
B41897
cellulase (EC 3.2.1.4) - Bacillus lautus
N:Alternate names: endo-1,4'-beta-glucanase
C:Species: Bacillus lautus
C:Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 15-Oct-1999
C:Accession: B41897; S27499
J:Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.
J. Bacteriol. 174, 3522-3531, 1992
A:Title: celsA from Bacillus lautus PU236 encodes a novel cellulose-binding endo-beta-
A:Reference number: A41897; MUID:92276330
A:Accession: B41897
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-700 <HAN>
A:Cross-references: EMBL:M76588; NID:g142661; PIDD:AAA22303.1; PID:g142663
A:Experimental source: PU236
A:Note: Sequence extracted from NCBI backbone (NCBIP:104605)
A:Function:

```



```

Db      423 A0IGASNVTFKFKLSSSVSGADYILEIGFKSAGQLOPGKDTGEIOMRPNKDMWSNYNQ 482
OY      120 TNDYSY-GTNTTFQDWTKTVYVNGVLWAGTE 150
        |||: : |||: : |||: : |||: : |||: :
Db      483 GNDMSWQMSWTSYGENEKVYAVIDGVLWAGOE 514

RESULT  11
S54744
cellulase (EC 3.2.1.4) CelV1 precursor - Erwinia carotovora (SCC 3193)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Erwinia carotovora
C:Variety: SCC 3193
C>Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999
C:Accession: S54744; S44996
R:Maer, A.; Heikimelmo, R.; Palva, E.T.
Mol. Gen. Genet. 247, 17-26, 1995
A:Title: Structure and regulation of the Erwinia carotovora subspecies carotovora SCC319
A:Reference number: S54744; MUID:95231512
A:Accession: S54744
A:Molecule type: DNA
A:Residues: 1-504 <MAE>
A:Cross-references: EMBL:X79241; NID:g493492; PIDN:CAAS5823.1; PID:g493493
C:Genetics:
A:Gene: celV1
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-33/Domain: signal sequence #status predicted <SIG>
F:33-504/Product: cellulase #status predicted <MAT>

Query Match      41.6%; Score 335.5; DB 2; Length 504;
Best Local Similarity 44.4%; Pred. No. 1,4e-22;
Matches 68; Conservative 31; Mismatches 47; Indels 7; Gaps 4;

OY      2 SGLKVOYKKNDSAPGDNQIKPGIQLVNTGSSVDSLTVYRYWFTPRDQ--GSSSTLYVNC 59
        |||: : |||: : |||: : |||: : |||: : |||: : |||: : |||: :
Db      354 TGDIVYLQYRNVDNPFDDAIRMAFNKNGSTPIKSLDQVRYFFHDDGRGANLTF--- 410
OY      60 DMAAMCGNIRASFGSVNPATPTADTYLQLSF-TGCTLAAGSTGEIQNRVKSMDSNFT 118
        |||: : |||: : |||: : |||: : |||: : |||: : |||: : |||: :
Db      411 DMAVNPNNIVSTGTGTPASTDKNRKYVLTFTFASGSLPGATGEVQVRIHAGDWSMNV 470
OY      119 ETNDYSYGTN-TTEQDQTKTVYVNGVLWAGTE 150
        |||: : |||: : |||: : |||: : |||: : |||: : |||: : |||: :
Db      471 ETNDYSYGPNTSYTNMCKITVADKGLTWAGTE 503

RESULT  12
A43802
cellulase (EC 3.2.1.4) / cellulose 1,4-beta-cellubiosidase (EC 3.2.1.91) - Caldocellum s
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Caldocellum saccharolyticum
C>Date: 30-Jan-1993 #sequence_revision 30-Sep-1993 #text_change 10-Jul-1998
C:Accession: A43802
R:Saul, D.J.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 56, 3117-3124, 1990
A:Title: celB, a gene coding for a bifunctional cellulase from the extreme thermophile
A:Reference number: A43802; MUID:91136262
A:Accession: A43802
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A:Molecule type: DNA
A:Residues: 1-915 <SAU>
A:Cross-references: EMBL:X13602
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
C:Keywords: cellulose degradation
C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>
```

```

Query Match      41.6%; Score 335; DB 2; Length 915;
Best Local Similarity 45.0%; Pred. No. 3e-22;
Matches 68; Conservative 29; Mismatches 50; Indels 4; Gaps 3;

OY      3 GGLKVOYKKNDSAPGDNQIKPGIQLVNTGSSVDSLTVYRYWFTPRDQSSSTLYVNC 62
        |||: : |||: : |||: : |||: : |||: : |||: : |||: : |||: :
Db      420 GQIKVLYANKETNSTYTITPMLKVVNNGSSSIDLSRVTIRYVYDGERAQAASV- 478
OY      63 AMCGNIRASFGSVNPATPTADTYLQLSFTGCT--LAAAGSTGEIQNRVKSMDSNFT 120
        |||: : |||: : |||: : |||: : |||: : |||: : |||: : |||: :
Db      479 QIGASNVTFKFKLSSSVSGADYILEIGFKSAGQLOPGKDTGEIQIRFKSDMSMNYNOG 538
OY      121 NDYSY-GTNTTFQDWTKTVYVNGVLWAGTE 150
        |||: : |||: : |||: : |||: : |||: : |||: : |||: : |||: :
Db      539 NDMSWQMSWTSYGENEKVYAVIDGVLWAGOE 569

RESULT  14
T17120
cellulase (EC 3.2.1.-) precursor, thermoactive - Caldocellum saccharolyticum
C:Species: Caldocellum saccharolyticum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C:Accession: T17120; A43745
R:Te'o, V.S.; Saul, D.J.; Bergquist, P.L.
Appl. Microbiol. Biotechnol. 43, 291-296, 1995
A:Title: CelA, another gene coding for a multidomain cellulase from the extreme therm
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:38:34 ; Search time 38.44 Seconds
(without alignments)
151.091 Million cell updates/sec

Title: US-09-917-384-5
Perfect score: 806
Sequence: 1 VSGGLKVGQYKKNDSAPGDNQ.....FQDMTKVTVVNGVLMGTE 150

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	380	47.1	145	1 YCEA_PAEIA	P29718 paenibacill
2	364	45.2	499	1 GUN2_BACSU	P10475 bacillus su
3	362	44.9	499	1 GUN3_BACSU	P23549 bacillus su
4	361.5	44.9	700	1 GUN4_PAEIA	P29719 paenibacill
5	359.5	44.6	772	1 CTRB_CLOTM	Q01866 clostridium
6	357.5	44.4	1853	1 CTRB_CLOTM	Q06851 clostridium
7	357	44.3	499	1 GUN1_BACSU	P07983 bacillus su
8	336	41.7	1331	1 MANB_CAUSA	P22533 caldocellum
9	335.5	41.6	504	1 GUNB_ERMCA	O59395 erwania car
10	335	41.6	1039	1 GUNB_ERMCA	P10474 c endogluc
11	335	41.6	1742	1 GUNB_ERMCA	P22534 caldocellum
12	325	40.3	505	1 GUN2_CLOSR	Q47036 erwania car
13	294.5	36.5	914	1 GUN2_CLOSR	P23659 clostridium
14	286	35.5	986	1 GUN1_CLOTM	Q02934 clostridium
15	253	31.4	879	1 GUN1_CLOTM	P38058 clostridium
16	228	28.3	1848	1 CBPA_CLOCL	O59394 erwania car
17	146.5	18.2	444	1 GUNB_ERMCA	P26225 cellulomona
18	112	13.5	1045	1 GUNB_ERMCA	P36693 rhizobium m
19	92.5	11.5	880	1 GUN4_ERMCA	P26221 thermomon
20	92.5	11.5	725	1 GUN4_ERMCA	P37700 clostridium
21	90.5	11.2	636	1 GUN4_ERMCA	P28622 bacillus sp
22	89.5	11.1	698	1 GUN4_ERMCA	P75882 escherichia
23	88.5	11.0	698	1 YWCA_ECOCI	P45797 paenibacill
24	81	10.0	238	1 GUNB_ERMCA	P11335 mumps virus
25	80.5	10.0	582	1 HEMA_MUMPR	O30611 pseudomonas
26	80.5	10.0	1148	1 ICERK_PSEBX	P81004 xenopus lae
27	80	9.9	282	1 ICERK_PSEBX	P06620 pseudobacill
28	79.5	9.9	1200	1 GUNB_ERMCA	P23904 pseudobacill
29	79	9.8	437	1 ICEN_PSEBX	Q47899 flavobacter
30	78.5	9.7	443	1 FLYS_FAME	O39163 wiseana iti
31	77.5	9.6	484	1 COAT_LRY9	P53316 methanococc
32	77	9.6	215	1 FTAI_METVA	P53328 aspergillus
33	76.5	9.5	211	1 XYNI_ASFPA	

34	76.5	9.5	211	1 XYNI_ASPTU	P55331 aspergillus
35	76.5	9.5	582	1 HEMA_MUMPR	P10866 mumps virus
36	76.5	9.5	1196	1 ICERK_PSEBX	O33479 pseudomonas
37	76	9.4	211	1 XYNI_ASFPA	P55329 aspergillus
38	76	9.4	282	1 FOR2_MELGA	P82013 melagris g
39	76	9.4	1157	1 GUNB_ERMCA	O45733 bacillus th
40	76	9.4	1210	1 ICEN_PSEBX	P09815 pseudomonas
41	75	9.3	247	1 FTAI_ERMCA	P57719 thermoplasma
42	74.5	9.2	211	1 XYNI_ASFPA	P33557 aspergillus
43	74.5	9.2	523	1 PUR9_MYCTU	P71553 m blifunctio
44	74	9.2	498	1 VNUC_IAUSS	P18073 influenza a
45	74	9.2	517	1 GUNB_ERMCA	P54937 clostridium

ALIGNMENTS

RESULT	1	STANDARD	PRT	145 AA
YCEA_PAEIA				
ID	P29718:			
AC	01-APR-1993 (Rel. 25, Created)			
DT	01-APR-1993 (Rel. 25, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Hypothetical protein in CELA 5' region (Fragment).			
OS	Paenibacillus lautus (Bacillus lautus).			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Paenibacillus.			
OX	NCBI_TaxID=1401;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PL236;			
RA	MEDLINE=92276330; PubMed=1592807;			
RX	Hansen C.K., Joergensen P.L., Diderichsen B.;			
RT	"celA from Bacillus lautus PL236 encodes a novel cellulose-binding			
FT	endo-beta-1,4-glucanase."			
RL	J. Bacteriol. 174:3522-3531(1992).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
DR	EMBL: M76588; AAA2302.1; ..			
DR	PIR: A41897; A41897.			
DR	PIR: S27498; S27498.			
DR	HSP: O06851; JNBC.			
DR	InterPro: IPR001956; CBD_3.			
DR	Pfam: PF00942; CBD_3; 1.			
DR	ProDom: PD001947; CBD_3; 1.			
KW	Hypothetical protein.			
FT	NON_TER			
FT	SEQUENCE 145 AA; 15782 MW; 9514E3A71B106AEB CRC64;			

Query Match 47.1%; Score 380; DB 1; Length 145;
Best Local Similarity 50.0%; Pred. No. 5.3e-28;
Matches 73; Conservative 29; Mismatches 38; Indels 6; Gaps 4;

OY	7 VQYKNDASDQKIRGQQLVNTGSSVDLTVYRYWFTSDGSSSTLYVNCDAAMGC 66	
DB	1 LQYAADTADNADNOIKPSFNKNGTSVADLTLKIRYFTDGSAAVNGW-IDNAOLGG 59	
OY	67 GNIRASGSVNPATPTADTYLQLSFT--GGTLAAGSTGEIONRNKNSMSTFTNTYS 124	
DB	60 SNIOISF--NHGTNSDITVELSSEAGSIAAGOSGETOLRMSKTIDWSFNENANDYS 117	
OY	125 Y-GTNTFQDMTKVTVVNGVLMGTE 149	
DB	118 PDGKTAFADMDRVLYVNGQIVMGTE 143	

```

RESULT 2
GUN2_BACSU STANDARD: PRT: 499 AA.
ID GUN2_BACSU
AC p10475;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Carboxymethyl-cellulase) (CMCase) (Cellulase).
GN BGIC OR GID OR EGIS.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
ON NCBI_TaxID=1423;
RX MEDLINE=87066783; PubMed=3024130;
RA MacKay R.M., Lo A., Willick G., Zuker M., Baird S., Dove M.,
RA Moranelli F., Selligy V.;
RA "Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene";
RA Nucleic Acids Res. 14:9159-9170(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CK-2;
RX MEDLINE=95225656; PubMed=7710280;
RA Lindahl V., Aa K., Tronamo A.;
RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus
RT subtilis CK-2.";
RL Antonie Van Leeuwenhoek 66:327-332(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Rose M., Entian K.;
RT Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE OF 30-45.
RC STRAIN=CK-2;
RX MEDLINE=95225655; PubMed=7710279;
RA Aa K., Flengsrud R., Lindahl V., Tronamo A.;
RT "Characterization of production and enzyme properties of an
RT endo-beta-1,4-glucanase from Bacillus subtilis CK-2 isolated from
RT compost soil.";
RL Antonie Van Leeuwenhoek 66:319-326(1994).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z29076; CAA82317.1; -
DR EMBL: X04689; CAA28392.1; -
DR EMBL: X67044; CAA47429.1; -
DR EMBL: Z73234; CAA97610.1; ALT_INIT.
DR EMBL: Z99113; CAB13696.1; ALT_INIT.
DR PIR: A26114; A26114.
DR HSSP: O85465; 1A3H.
DR Subtilist; BG10437; bg1C.
DR Interpro: IPR001956; CBD_3.
DR Interpro: IPR001547; Glyco_hydro_F5.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00150; cellulase; 1.
DR ProDom: PD001947; CBD_3; 1.
DR ProSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.

```

```

KW Cellulose degradation: Hydrolase; Glycosidase; Signal;
KM Complete proteome.
FT SIGNAL 1 29
FT CHAIN 30 499
FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
FT CONFLICT 283 283 S -> N (IN REF. 2).
SQ SEQUENCE 499 AA; 55287 MW; 8F735FE711B3EAE2 CRC64;

Query Match 45.2%; Score 364; DB 1; Length 499;
Best Local Similarity 44.6%; Pred No. 6,5e-26;
Matches 66; Conservative 32; Mismatches 46; Indels 4; Gaps 2;

QY 4 GLKQYKRNDSAPGDNQIRPGQLVNTGSSVDLSTVRYWF-TRDGSSTLYVNCDA 62
D 353 GISVQYRAGDGSWNSNQIRPOLIKNGNWTYDLKVTARYWKANKQGN--FDCDYA 409
QY 63 AMCCGNIRASFGSVNPATPTADTYLQSLFTGGLAAGSGTGEIQNRVKSMSNPTETMD 122
D 410 QIGCGNVTHTKFTYLRKPGADTYLDELGRKNGTLAPGASTGNIQRLHNDWMSNVAQSGD 469
QY 123 YSYGTNFTFQDMTKRYTVYVNGVLWGTE 150
D 470 YSFRKSNTRFTTKITTYDQGLIWGTE 497

RESULT 3
GUN3_BACSU STANDARD: PRT: 499 AA.
ID GUN3_BACSU
AC P23549;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Carboxymethyl-cellulase) (CMCase) (Cellulase).
GN BGIC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
ON NCBI_TaxID=1423;
RX MEDLINE=91299280; PubMed=1368694;
RA Park S.H., Kim H.K., Pack M.Y.;
RT "Characterization and structure of the cellulase gene of Bacillus
RT subtilis BSE616.";
RL Agric. Biol. Chem. 55:441-448(1991).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D01057; BAA00859.1; -
DR PIR: JN0111; JN0111.
DR HSSP: O85465; 1A3H.
DR Interpro: IPR001956; CBD_3.
DR Interpro: IPR001547; Glyco_hydro_F5.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00150; cellulase; 1.
DR ProDom: PD001947; CBD_3; 1.
DR ProSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KM Cellulose degradation: Hydrolase; Glycosidase; Signal.

```


	Matches	70;	Conservative	34;	Mismatches	44;	Indels	11;	Gaps	3;
Oy	1	VSGGLVYQKKNDSAPCGDNIQIRPGQLVNTGSSSYDLSVTARYWFTFRGGSSLIYNCD	60							
Db	365	VSGLKEEFLNSPSTLTMSINPQEFVNTHSSAIDLRLRYLVYVDGGRDQTFW-CD	423							
Oy	61	WAAM-----GCCNIRASFGSVNPATPTADTYLQISFETGTLAGGSTGEIQNRVK	111							
Db	424	HAAIIGSGSYNGTTSNVKOTFKVMSSSTNNMDTYLEISFTGLEPAHW-OIOGRAPK	482							
Oy	112	SDMSEFTETNDYSXGTLTTTEQDTAKVTVYVNGYLWGCE	150							
Db	483	NDMSNYTQSNDYSPKSAQSEVEEDVQYTAALINGVLWGKE	521							

DR	PIR; A26874; A26874.
DR	HSSP; O85465; 1A3H.
DR	Interpro; IPR001956; CBD_3.
DR	Interpro; IPR001347; Glyco_hydro_F5.
DR	Pfam; PF00942; CBD_3; 1.
DR	Pfam; PF00150; cellulase; 1.
DR	Prodom; PD001947; CBD_3; 1.
DR	ProSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR	Cellulose degradation; Hydrolase; Glycosidase; Signal.
KW	1
FT	SIGNAL
FT	CHAIN
FT	ACT_SITE
FT	ACT_SITE
FT	DOMAIN
SQ	SEQUENCE

```

Db 353 GVSQYKAGGRVNSQIRPOLHIKNNGNAVLDKVTARYTNVANKNGON--EDCOYA 409
OY 63 AMCGGNIRAFSGSVNPAFPTADTVLQLSFTFGTLAGSGTGEIONRYKNSDMSNFETEND 122
| | | | | : : : : | | | | | : | | | | | : | | | | : : : :
Db 410 QMGCGNLTFRFVTLHHRKQCADTVLELFGPKGTSLSPGASGTGNIOLRINDMSNTAOSGD 469
OY 123 YSYGNTTFQDMTRKYVYVNGVLTWGT 150
| | : | | : | | | | | | | | | |
Db 470 XSFQSNFETKTKITLYLRQGLWGT 497

RESULT 8
MANB-CALSA STANDARD: PRT: 1331 AA.
ID P22533:
AC 01-AUG-1991 (Rel. 19, Created)
AD 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE beta-mannanase/endorucanase A precursor [includes: Mannan endo-1,4-
beta-mannosidase A (EC 3.2.1.78) (beta-mannanase) (Endo-1,4-
mannanase); Endo-1,4-beta-glucanase (EC 3.2.1.4) (Cellulase)].
GN MANA.
OS Caldocellum saccharolyticum (Caldicellulosigrupor saccharolyticus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermanaerobacter group; Caldicellulosigrupor.
OX NCBI_Taxid=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93119139; PubMed=1476429;
RA Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;
RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a
RT multidomain enzyme.";
RL Appl. Environ. Microbiol. 58:3864-3867(1992).
RN [2]
RP SEQUENCE OF 1-346 FROM N.A.
RX MEDLINE=91247819; PubMed=2039230;
RA Luehl E., Jasmet N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a
RT gene coding for a beta-mannanase from the extremely thermophilic
RT bacterium 'Caldocellum saccharolyticum'.";
RL Appl. Environ. Microbiol. 57:694-700(1991).
CC -1- FUNCTION: DEGRADATION OF HEMICELLULOSES. THE SECOND MOST ABUNDANT
CC POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH
CC MANNANASE AND ENDOSUGCANASE ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
CC linkages in mannans, galactomannans, glucamannans, and
CC galactoglucamannans.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- MISCELLANEOUS: THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREES
CC CELSIUS.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC J (FAMILY 44 OF GLYCOSYL HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on way
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/).
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L0157; AAA71887.1; -
CC EMBL: M36063; AAA72861.1; -
CC PIR: B43745; B43745.
CC PIR: A48954; A48954.
CC HSSP: 006851; INBC.
CC InterPro: IPR001956; CBD_3.
CC InterPro: IPR001547; Glyco_hydro_F5.
CC Pfam: PF00942; CBD_3; 2.
CC Pfam: PF00150; cellulase; 1.
CC

```

DR PRODOM: PD001947; CBD_3; 2
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KM Hydrolyase; Glycosidase; Cellulose degradation; Signal;
 FT Multifunctional enzyme.
 FT SIGNAL 1 41
 FT CHAIN 42 1331
 FT DOMAIN 42 325
 FT DOMAIN 326 361
 FT DOMAIN 362 518
 FT DOMAIN 519 564
 FT DOMAIN 565 720
 FT DOMAIN 721 780
 FT DOMAIN 781 1331
 FT ACT_SITE 162 162
 FT ACT_SITE 257 257
 FT CONFLICT 338 338
 FT CONFLICT 340 346
 FT SEQUENCE 1331 AA; 146892 MW; FFBCA51BB8DF0E0 CRC64;

Query Match 41.7%; Score 336; DB 1; Length 1331;
 Best Local Similarity 44.7%; Pred. No. 7.2e-23;
 Matches 68; Conservative 29; Mismatches 51; Indels 4; Gaps 3;

OY 2 SGGGLVQYKNNDSAPGDNQIKFGLQLVNTGSSVDLSTVTVYKFTROGSGSTLYVNCW 61
 DB 364 SQIIVLVANKETSTNTIRFWLVNVSSTIDSLRVTIRYTWVDERAQSALS-DW 422
 OY 62 AAMGCGNIRASFVSNPATPTADTYQLSF-TGGLAAGSGTEIONRVKSDMSNFT 119
 DB 423 AQIGASNTYFKFVKLSSVSAGDYILEIGFKSGAGLOPGCKRTGELQMFNNDMSNYNQ 482
 OY 120 TNDYST-GTNTFDQMTKYTVYVNGVLWGTE 150
 DB 483 GNDMSWISQMTSYGENEKVTAYIDGLVWGQE 514

RESULT 9
 GUNB_ERMCA STANDARD; PRT; 504 AA.
 ID GUNB_ERMCA
 AC 059395;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Endoglucanase VI precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase VI)
 DE (Cellulase VI).
 GN CELV1.
 OS Erwina carotovora.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Pectobacterium.
 OX NCBI_TaxID=554;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SCC3193;
 RX MEDLINE=95231512; PubMed=7715600;
 RA Mae A., Heikinhelmo R., Palva E.T.;
 RT "Structure and regulation of the Erwina carotovora subspecies
 RT carotovora SCC3193 cellulase gene celv1 and the role of cellulase in
 RT phytopathogenicity";
 RL Mol. Gen. Genet. 247:17-26(1995).
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC EMBL; X79241; CAA55823.1; -.
 DR HSSP; 085465; 1A3H.
 DR InterPro; IPR001956; CBD_3.
 DR InterPro; IPR001547; Glyco_hydro_F5.
 DR Pfam; PF00942; CBD_3; 1.
 DR Pfam; PF00150; cellulase; 1.
 DR PRODOM: PD001947; CBD_3; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KM Cellulose degradation; Hydrolyase; Glycosidase; Signal;
 FT SIGNAL 1 31
 FT CHAIN 32 504
 FT DOMAIN 32 334
 FT DOMAIN 335 352
 FT DOMAIN 353 504
 FT ACT_SITE 168 168
 FT ACT_SITE 256 256
 FT SEQUENCE 504 AA; 54963 MW; 0D7EC747815655FA CRC64;

Query Match 41.6%; Score 335.5; DB 1; Length 504;
 Best Local Similarity 44.4%; Pred. No. 2.6e-23;
 Matches 68; Conservative 31; Mismatches 47; Indels 7; Gaps 4;

OY 2 SGGGLVQYKNNDSAPGDNQIKFGLQLVNTGSSVDLSTVTVYKFTROG--GSSTLYVNC 59
 DB 354 TGDIVQYKNNVNSDDAIRAFNFKNTGSPRIKSDIQYTFYFHDGKPGANLFV--- 410
 OY 60 DWAAMCGNIRASFVSNPATPTADTYQLSF-TGGLAAGSGTEIONRVKSDMSNFT 118
 DB 411 DWANVPNNIVYSTGTTPASTKARIVLYTVASGSLQPGAEFGVVRHAGDMSNVN 470
 OY 119 ETNDYSYGTN-TTFQDMTKYTVYVNGVLWGTE 150
 DB 471 ETNDYSYGNITSYTNMWDKITYHDKGLTWGTE 503

RESULT 10
 GUNB_CALSA STANDARD; PRT; 1039 AA.
 ID GUNB_CALSA
 AC P10474;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Endoglucanase/exoglucanase B precursor [includes: Endoglucanase
 DE (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase)
 DE (Cellulohydrolyase); Exoglucanase (EC 3.2.1.91) (Exocellulohydrolyase)]
 DE (1,4-beta-cellobiohydrolyase)].
 GN CELB.
 OS Caldocellum saccharolyticum (Caldicellulosigruppor saccharolyticus).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Thermoaerobacter group; Caldicellulosigruppor.
 OX NCBI_TaxID=44001;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89098398; PubMed=2789517;
 RA Saul D.J., Williams L.C., Love D.R., Chamley I.W., Bergquist P.I.;
 RT "Nucleotide sequence of a gene from Caldocellum saccharolyticum
 RT encoding for exoglucanase and endoglucanase activity";
 RL Nucleic Acids Res. 17:439-439(1989)
 CC -1- FUNCTION: THIS PROTEIN IS MADE UP OF TWO DOMAINS: THE N-TERMINAL
 CC DOMAIN HAS EXOGLUCANASE ACTIVITY WHILE THE C-TERMINAL DOMAIN IS
 CC AN ENDOGLUCANASE.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellobioetraose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
 CC F (FAMILY 10 OF GLYCOSYL HYDROLASES).
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
 CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL: X13602; CAA31936.1; -
 PIR: S02711; S02711.
 HSSP: Q06851; INBC.
 DR Interpro: IPR001956; CBD_3.
 DR Interpro: IPR001000; Glyco_hydro.10.
 DR Interpro: IPR001547; Glyco_hydro.P5.
 DR Pfam: PF00942; CBD_3; 1.
 DR Pfam: PF00150; cellulase; 1.
 DR Pfam: PF00331; Glyco_hydro.10; 1.
 DR PRINTS: PR00134; GLYDRLASE10.
 DR PRODOM: PD001947; CBD_3; 1.
 DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR Cellulose degradation; Hydrolase; Glycosidase; Repeat;
 KW Multifunctional enzyme; Signal.
 FT SIGNAL 1 28
 FT CHAIN 29 1039
 FT DOMAIN 376 416
 FT DOMAIN 417 570
 FT DOMAIN 571 618
 FT ACT_SITE 177 177
 FT ACT_SITE 285 285
 FT ACT_SITE 792 792
 FT ACT_SITE 1039 AA; 117641 MW; 0E0378171594DDAE CRC64;
 SQ SEQUENCE

Query Match 41.6%; Score 335; DB 1; Length 1039;
 Best Local Similarity 45.0%; Pred. No. 6.7e-23;
 Matches 68; Conservative 29; Mismatches 50; Indels 4; Gaps 3;

OY 3 GGLKVOYKNNDSAPGDNQIKPGIQLVNTGSSVDLSTVYRYWFTPRDGGSSLTLYNCMA 62
 DB 420 GGLKVLANKETSTNTTIRPMLKVNSSSSSIDLSRTYIRKTYVDDERASANS-DMA 478
 OY 63 AMGCGNIRAFSGSVNPAITADTYLQSLFTGCT--LAAGSTGEIONRVKNSDMSNFTET 120
 DB 479 QIGASNVTFKFKLVSSVSADYLYLEIFKSGAGOLPCKDVGELQIRFNKSDMSNINOG 538
 OY 121 NDYSY-GTNTTFQDMTKRVTVYVNGVLWGTET 150
 DB 539 NDMSWLSMTSYGENEKRYTAYIDGVLWGOE 569

RESULT 11
 GUNA_CALSA STANDARD; PRT; 1742 AA.
 AC P22534;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)
 DE (Cellulase A).
 GN Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
 OS Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Thermoaerobacter group; Caldicellulosiruptor.
 NCBI_TaxID=44001;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95336703; PubMed=7612247;
 RA Te'O V.S., Saul D.J., Bergquist P.L.;
 RT "CellA, another gene coding for a multidomain cellulase from the
 extreme thermophile Caldocellum saccharolyticum."
 RT Appl. Microbiol. Biotechnol. 43:291-296(1995).
 RN [2]

SEQUENCE OF 1516-1742 FROM N.A.
 RX MEDLINE=91247819; PubMed=2039230;
 RA Leuthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
 RT "Cloning, sequence analysis, and expression in Escherichia coli of a
 gene coding for a beta-mannanase from the extremely thermophilic
 bacterium 'Caldocellum saccharolyticum'";
 RT Appl. Environ. Microbiol. 57:694-700(1991).
 CC -1- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
 ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL
 DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
 CELLULOSE.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 linkages in cellulose. (ALSO TERMED "HINGE") MAY BE A POTENTIAL
 CC -1- PTM: THE LINKER REGION (ALSO TERMED "HINGE") BELONGS TO CELLULOSE FAMILY
 SITE FOR PROTEOLYSIS.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
 E (FAMILY 9 OF GLYCOSYL_HYDROLASES).
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
 L (FAMILY 48 OF GLYCOSYL_HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

EMBL: L32742; AAA91086.1; -
 EMBL: M36063; AAA72860.1; -
 EMBL: L01257; -; NOT_ANNOTATED_CDS.
 PIR: A43745; A43745.
 DR PIR: A43745; A43745.
 DR HSSP: P26221; IPR4.
 DR Interpro: IPR001956; CBD_3.
 DR Interpro: IPR000556; Glyco_hydro.48.
 DR Interpro: IPR001701; Glyco_hydro.9.
 DR Pfam: PF00942; CBD_3; 3.
 DR Pfam: PF02011; Glyco_hydro.48; 1.
 DR Pfam: PF00759; Glyco_hydro.9; 1.
 DR PRINTS: PR00844; GLYDRLASE48.
 DR PRODOM: PD001947; CBD_3; 2.
 DR PRODOM: PD01903; Glyco_hydro.48; 1.
 DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
 DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
 DR Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
 FT SIGNAL 1 23
 FT CHAIN 24 1742
 FT DOMAIN 24 642
 FT DOMAIN 643 700
 FT DOMAIN 701 857
 FT DOMAIN 858 903
 FT DOMAIN 904 1060
 FT DOMAIN 1061 1112
 FT DOMAIN 1113 1742
 FT ACT_SITE 396 396
 FT ACT_SITE 434 434
 FT ACT_SITE 443 443
 FT CONFLICT 1545 1545
 FT CONFLICT 1742 AA; 193696 MW; 3F0699A2123BED07 CRC64;
 SQ SEQUENCE

Query Match 41.6%; Score 335; DB 1; Length 1742;
 Best Local Similarity 44.7%; Pred. No. 1.2e-22;
 Matches 68; Conservative 29; Mismatches 51; Indels 4; Gaps 3;

OY 2 SGLKVOYKNNDSAPGDNQIKPGIQLVNTGSSVDLSTVYRYWFTPRDGGSSLTLYNCMA 61
 DB 704 SGLKVLANKETSTNTTIRPMLKVNSSSSSIDLSRTYIRKTYVDDERASANS-DW 762
 OY 62 AMGCGNIRAFSGSVNPAITADTYLQSLFTGCT--LAAGSTGEIONRVKNSDMSNFTET 119
 DB 763 AIGASNVTFKFKLVSSVSADYLYLEIFKSGAGOLPCKDVGELQIRFNKSDMSNINO 822

QY 61 MAAGCGNIRASFGSVNPTATDYTLQSLFT--GGTLAAGSGTGEIONRVKNSDMSNFT 118
 DB 822 MASIGSSNVTGTFVKMDGATGTGADVLYLEIGFPQAGTLEPGAST-EVQGRSKIDMDYDT 880
 QY 119 ETNDYSYSG--TNTFQDMTKVTVYVNGVLWGTE 150
 DB 881 QTNDSYFNPTASSYVDNKITATVYISGNLYGIE 913

RESULT 14

GUNI_CLOSR STANDARD: PRT: 986 AA.

AC P23659;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Endoglucanase 2 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Thermococcus cellulase) (Avicelase I).
 GN CELZ.
 OS Clostridium stercorarium.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OC NCBI_TaxID=1510;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 26-36 AND 475-486.
 RC STRAIN-NCIB 11745;
 RX MEDLINE-91066838; PubMed-2250652;
 RA Jauris S., Ruecknagel K.P., Schwarz W.H., Kratzsch P.,
 RA Bronnenmeier K., Staudenbauer W.L.;
 RA "Sequence analysis of the Clostridium stercorarium celz gene encoding
 RT a thermocatalytic cellulase (Avicelase I): Identification of catalytic
 RT and cellulose-binding domains.";
 RL Mol. Gen. Genet. 223:258-267 (1990).
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- SIMILARITY: CONTAINS MULTIPLE CELLULOSE-BINDING SITES.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY E (FAMILY 9 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X55299; CAA39010.1; AUL-SEQ.
 DR PIR: S12021; S12021.
 DR HSSP: P26221; 1TF4.
 DR InterPro: IPR001956; CBD_3.
 DR InterPro: IPR001701; Glyco_hydro_9.
 DR Pfam: PF00942; CBD_3; 2.
 DR Pfam: PF00759; Glyco_hydro_9; 1.
 DR ProDom: PD001947; CBD_3; 1.
 DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_2; 1.
 DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
 DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
 FT SIGNAL 1 25
 FT CHAIN 1 25
 FT REPEAT 26 986 ENDOGLUCANASE 2.
 FT REPEAT 498 641
 FT REPEAT 651 738 DOMAIN C'.
 FT REPEAT 744 831 DOMAIN B'.
 FT REPEAT 854 986 DOMAIN C'.
 FT DOMAIN 835 986 CELULOSE-BINDING (BY SIMILARITY).
 FT ACT_SITE 400 400 BY SIMILARITY.
 FT ACT_SITE 438 438 BY SIMILARITY.
 FT ACT_SITE 447 447 BY SIMILARITY.
 FT ACT_SITE 986 AA; 109512 MW; 180280982293690 CRC64;
 SQ SEQUENCE

Query Match 35.5%; Score 286; DB 1; Length 986;

Best Local Similarity 37.9%; Pred. No. 1,8e-18; Indels 8; Gaps 3;
 Matches 58; Conservative 36; Mismatches 51;

QY 2 SGLKAYQVKNDSACPDNDIKRGDQVNTGSSSVYTVYVWFTRDGSSTLVNCDW 61
 DB 837 TGVQIOFMNGNTSDKTNKIMPRITLNTGTPILSDVKIRYITIDGKDNQW-CDW 895
 QY 62 MAAGCGNIRASFGSVNPTATDYTLQSLFTGTLAAG---GSTGEIQNRVNSDMSNFT 117
 DB 896 SSVGSNNITCTFFKMAPEKRGADVYLETGFTDG---AGLQPNOSIEVQNRKADMDY 952
 QY 118 ETNDYSYSGTNTFQDMTKVTVYVNGVLWGTE 150
 DB 953 IQTNDSYFNPTASSYVDNKITATVYISGNLYGIE 985

RESULT 15

GUNI_CLOTM STANDARD: PRT: 879 AA.

AC Q02934;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Endoglucanase I precursor (EC 3.2.1.4) (Egi) (Endo-1,4-beta-glucanase)
 DE (Cellulase I).
 GN CELL.
 OS Clostridium thermocellum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OC NCBI_TaxID=1515;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 56-69.
 RC STRAIN-NCIB 10682;
 RX MEDLINE-93171873; PubMed-8436949;
 RA Hazlewood G.P., Davidson K., Laurie J.I., Huskisson N.S.,
 RA Gilbert H.J.;
 RA "Gene sequence and properties of Cell, a family E endoglucanase from
 RT Clostridium thermocellum.";
 RL J. Gen. Microbiol. 139:307-316 (1993).
 CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-D-
 CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
 CC GLUCANS. PRINCIPALLY ACTIVE AGAINST BARLEY BETA-GLUCAN.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- PATHWAY: CELLULOSE DEGRADATION.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY E (FAMILY 9 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U04735; AAA20892.1; -.
 DR PIR: A47704; A47704.
 DR HSSP: P26221; 1TF4.
 DR InterPro: IPR001956; CBD_3.
 DR InterPro: IPR001701; Glyco_hydro_9.
 DR Pfam: PF00942; CBD_3; 2.
 DR Pfam: PF00759; Glyco_hydro_9; 1.
 DR ProDom: PD001947; CBD_3; 1.
 DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_2; 1.
 DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
 DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 55
 FT CHAIN 56 879
 FT DOMAIN 56 879 CATALYTIC.
 FT DOMAIN 729 879 CELULOSE-BINDING (BY SIMILARITY).
 FT ACT_SITE 448 448 BY SIMILARITY.
 FT ACT_SITE 486 486 BY SIMILARITY.

FT	ACT_SITE	495	495	BY SIMILARITY.
SQ	SEQUENCE	879 AA;	97796 MW;	35A60069A514A927 CRC64;

Query Match	31.4%	Score 253;	DB 1;	Length 879;
Best Local Similarity	37.0%	Pred. No. 1.6e-15;		
Matches 47; Conservative	34;	Mismatches 42;	Indels 4;	Gaps 3;

```

0y      1 VSGGLKYYVKNNDSAPGDNQIKPGIOLYNNQSSVDLSITVYRWFPRDDGSSSTLYVNC 60
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      736 IKEGVVLLDYANGNMAQVTSINPFRKIIINQTKMINLSDVKIRYYRKEGASQNFV-CD 794
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

0y      61 WAAAGCGNIRASFSGSVNPATPDTADTYIOLSF--TGTTLAAGSGTGEIONVKSQDSMFT 118
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      735 WSSAGNSVNTGCFNFMLSPKKGADPTCLEFVGFGAGGATLDPEGSV-EVQIRPSKEDWSNYN 853
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

0y      119 ETNDYSY 125
        ::|||::
Db      854 QSDNYSF 860

```

Search completed: August 29, 2002, 16:38:35
Job time: 955 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: August 29, 2002, 16:37:45 ; Search time 125.36 Seconds

(Without alignments)
206.998 Million cell updates/sec

Title: US-09-917-384-5

Perfect score: 1 VSGGLKVOYKNNDSAPGDNQ.....FQDWTKVTVYVNGVLWGTE 150

Sequence: 1 VSGGLKVOYKNNDSAPGDNQ.....FQDWTKVTVYVNGVLWGTE 150

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000
Maximum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_19: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mmc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_prodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriophage: *
17: sp_archaeal: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	434	53.8	170	2	Q9RFK6	Q9RFK6 caldicellul
2	430	53.3	921	2	Q9RLB8	Q9RLB8 caldicellul
3	426	52.9	930	2	Q9RFK5	Q9RFK5 caldicellul
4	371	46.0	997	2	Q924I1	Q924I1 bacillus su
5	364	45.2	499	2	Q937U6	Q937U6 bacillus su
6	364	45.2	508	2	Q93LD0	Q93LD0 bacillus su
7	357	44.3	499	2	Q52731	Q52731 bacillus su
8	355	44.0	499	2	Q45532	Q45532 bacillus su
9	350	43.4	501	2	Q83012	Q83012 bacillus su
10	348.5	43.2	1711	2	P96311	P96311 anaerocellu
11	345	42.8	1000	2	Q24820	Q24820 thermophil
12	343	42.6	1770	2	Q9X3P5	Q9X3P5 caldicellul
13	341	42.3	1779	2	Q52374	Q52374 caldicellul
14	340	42.2	996	2	Q9A0H0	Q9A0H0 caldicellul
15	339	42.1	1751	2	Q9A0G4	Q9A0G4 caldicellul
16	336	41.7	1426	2	Q9X3P6	Q9X3P6 caldicellul

17	316	39.2	261	2	Q9A0G7	Q9A0G7 caldicellul
18	315.5	39.1	1915	2	Q9RPL0	Q9RPL0 actividrio
19	315	39.1	486	2	Q45430	Q45430 bacillus sp
20	307.5	38.2	473	2	Q9RK75	Q9RK75 streptomyce
21	303.5	37.7	887	2	Q9L3J8	Q9L3J8 clostridium
22	284	35.2	2316	2	Q9FDJ9	Q9FDJ9 bacteroides
23	282.5	35.0	1483	16	Q977Y4	Q977Y4 clostridium
24	270	33.5	1162	2	Q82830	Q82830 clostridium
25	265	32.9	1546	2	Q45996	Q45996 clostridium
26	254.5	31.6	1230	2	Q59325	Q59325 clostridium
27	252.5	31.3	92	2	Q9L8L9	Q9L8L9 caldicellul
28	206	25.6	321	2	Q46392	Q46392 clostridium
29	182.5	22.6	660	2	Q9L3J2	Q9L3J2 clostridium
30	129.5	16.1	202	2	Q9A0G6	Q9A0G6 caldicellul
31	123.5	15.3	197	2	Q9A0G1	Q9A0G1 caldicellul
32	98	12.2	715	2	Q65987	Q65987 clostridium
33	89.5	11.1	659	2	Q9R9H6	Q9R9H6 bacillus pu
34	87.5	10.9	319	3	P78869	P78869 achlosacch
35	87.5	10.9	712	16	Q97RK5	Q97RK5 clostridium
36	86	10.7	440	2	Q93L19	Q93L19 bacillus sp
37	85	10.5	428	3	Q12646	Q12646 neocallimas
38	85	10.5	582	16	P71868	P71868 mycobacteri
39	84.5	10.5	582	12	Q83345	Q83345 mumps virus
40	84	10.4	781	2	Q9S261	Q9S261 streptomyce
41	83.5	10.4	359	5	Q95WD5	Q95WD5 pristina le
42	83	10.3	936	2	Q93IE6	Q93IE6 ruminooccu
43	82.5	10.2	257	13	Q90536	Q90536 ginglymocto
44	82.5	10.2	438	10	Q9FHB7	Q9FHB7 arbidopsia
45	82.5	10.2	582	12	Q83347	Q83347 mumps virus

ALIGNMENTS

RESULT 1
ID Q9RFK6 PRELIMINARY; PRT: 170 AA.
AC Q9RFK6;
DT 01-MAY-2000 (TREMREL.13, Created)
DT 01-MAY-2000 (TREMREL.13, Last sequence update)
DT 01-DEC-2001 (TREMREL.19, Last annotation update)
DE HYPOTHETICAL 18.5 KDA PROTEIN (FRAGMENT).
OS Caldicellulius cellulosovorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Alicyclobacillus group; Caldicellulius.
OX NCBI_TaxID=74586;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10653733;
RA Suna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
RT Caldicellulius cellulosovorans and action of the recombinant enzyme on
RT kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670(2000).
DR EMBL: AF163837; AAF22273.1; -.
DR HSSP: Q06851; INBC.
DR InterPro: IPR001956; CBD_3.
DR Pfam: PF00942; CBD_3; 1.
DR Prodom: PD001947; CBD_3; 1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 170 AA; 18493 MW; 7AC9D33F44E3A0B4 CRC64;

Query Match 53.8%; Score 434; DB 2; Length 170;
Best Local Similarity 52.3%; Pred. No. 1.2e-29;
Matches 79; Conservative 29; Mismatches 39; Indels 4; Gaps 3;

QY 3 GGLVQYKNNDSAPGDNQIPGLQVNTGSSVVDISTVTVRWYFTROGSSLTLYVNCMA 62
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 20 GSLVQYRADPTNAGDNQKLFKFRVNRGTSSVPLSETITRYWTVVD-GDPQVFNCDMA 78
QY 63 AMGGNIRASFGSVNPAITPRADTYLQLSFT--GGTLAAGSGTGEIQNRVKSMDMSNFTET 120

[illegible][illegible]

50	SEQUENCE	930 AA;	101576 MW;	0086638D54D1A2CC	CRC64;
DR	PROSITE;	PS00659;	GLYCOSYL_HYDROL_F5;	UNKNOWN_1.	
DR	PRINTS;	PF03067;	Chitin_bind_3; 1.		
DR	PFam;	PF00150;	cellulase; 1.		
DR	InterPro;	IPR001956;	CBD_3.		
DR	InterPro;	IPR004302;	Chitin_bind_3.		
DR	InterPro;	IPR001547;	Glyco_hydr_F5.		
DR	InterPro;	IPR002965;	P_fltch_extensn.		
DR	Pfam;	PF00942;	CBD_3; 2.		
DR	Pfam;	PF00150;	cellulase; 1.		
DR	PRINTS;	PF03067;	Chitin_bind_3; 1.		
DR	ProDom;	PD001947;	CBD_3; 2.		
DR	PROSITE;	PS00659;	GLYCOSYL_HYDROL_F5;	UNKNOWN_1.	
DR	Signal.				
FT	SIGNAL	1	33	POTENTIAL.	
FT	CHAIN	34	930	MULTIDOMAIN_BETA-1,4-MANNANASE.	
50	SEQUENCE	930 AA;	101576 MW;	0086638D54D1A2CC	CRC64;

```

Query Match      52.9%   Score 426; DB 2; Length 930;
Best Local Similarity    52.0%; Pred. No. 4,1e-28;
Matches 78; Conservative 35; Mismatches 35; Indels 2; Gaps 2;

QY      2 SGGKLVQYRKNNDSAPGDNOIKPGIQLVNTGSSSYDLSTVTVRVFTFRDGSSTLVNCDW 61
        |||::||: : |::||::||::||: : ||::||: : |||
Db       287 SGLTVEVERVDFTSATDNOMKPLRIYVTVSGOAVDLFELKRVRYMTKN-STQAQDYPCDW 345
        |||::||| |::||: : ||::||::||: : ||::||: : |||

QY      62 AAMCGGNTRASEGVSVPNRPTPADRYLOLSPFGCTLAAGSGEIQONRNKSDMSNFETTN 121
        ||::||| |::||: : ||::||::||: : ||::||: : |||
Db       346 AQTGGSNTRAFVSLISQPVSGADSTLETFSGSIPACNGTGELIONRHHFTMMNYETD 405
        ||::||| |::||: : ||::||::||: : ||::||: : |||

QY      122 DYSY-GINTFEODMKTIVTVYNGVLVMGTSE 150
        ||::|| |::||: : ||::||::||: : ||::||: : |||
Db       406 DWSYGAAQTWGPSTRTITLRNGVLMGTSE 435
        ||::|| |::||: : ||::||::||: : ||::||: : |||

RESULT      4
Q9Z4T1      PRELIMINARY; PRT; 997 AA.
AC          Q9Z4T1;
DT      01-MAY-1999 (TREMBLrel. 10, Created)
DT      01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      CELLULASE PRECURSOR (EC 3.2.1.4).
GN      CELB.
OS      Bacillus sp. Bp-23.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group;
OC      Bacillus/Staphylococcus group; Bacillus.
OX      NCBI_TaxID=89769;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-BP-23.
RX      MEDLINE-21129642; PubMed-11234960;
RA      Pastor F.I.J., Puljo X., Blanco A., Vidal T., Torres A.L., Diaz P.;
RT      "Molecular cloning and characterization of a multidomain endoglucanase
RT      from Paenibacillus sp Bp-23: evaluation of its performance in pulp
RT      refining.";
RL      Appl. Microbiol. Biotechnol. 55:61-68(2001).
CC      -!- SIMILARITY: COMPAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
DR      EMBL: AJ133614; CAB38941.1; -.
DR      HSSP; P26221; ITFA.
DR      InterPro; IPR001956; CBD_3.
DR      InterPro; IPR003961; FN_III.

```

Tue, Sep 3 13:42:05 2002

us-09-917-384-5.rspt

DR InterPro: IPR001701; Glyco_hydro_9.
 DR InterPro: IPR000566; Lipocalin_cytrFABP.
 DR Pfam: PF00942; CBD_3; 2.
 DR Pfam: PF00041; fn3; 2.
 DR Pfam: PF00759; Glyco_hydro_9; 1.
 DR ProDom: PD001947; CBD_3; 1.
 DR ProDom: PD00060; FN3; 2.
 DR SMART: SM000592; GLYCOSYL_HYDROL_F9_1; 1.
 DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
 DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
 DR GlycoSite: Hydrolyase; Signal.
 KW SIGNAL 1 34
 FT SEQUENCE 997 AA; 106927 MW; F20CB119D6410629 CRC64;

Query Match 46.0%; Score 371; DB 2; Length 997;
 Best Local Similarity 47.4%; Pred. No. 2.1e-23;
 Matches 72; Conservative 28; Mismatches 48; Indels 4; Gaps 3;

OY 2 SGLKVOYKNNDSAPGDNQIKPGLQVNTGSSVDLSVTYVRYWFTRDGGSTLYVNCDA 61
 DB 846 TGLEVQYRSGSGNSNAVYPOFNLKNTGQALDLSVKIRYFTKD-GTEELSEWCXY 904
 OY 62 AMCGCNINAFSGSVNPAFTADTYLQLSFT--GGTLAAGSGTEIQNRNKSMDMSFETND 119
 DB 905 AQVGSANVQGMVAVNPAKGTADTYVEISFTSGAGSLAAGETGVIOTRFSKNMSAFDQ 964
 OY 120 TNDYSY-GTNTTFQDMTKVTVVNGVLVWGTE 150
 DB 965 SNDYSYASKTAFANMKYVAVOGNTQVWGLE 996

RESULT 5
 O93JTV6 PRELIMINARY; PRT; 499 AA.
 ID O93JTV6
 AC 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 RT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE ALKALI TOLERABLE CELLULASE.
 GN CEL.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Y106;
 RA Q1 M., Wang P., Liu X., Qu Y.;
 RT "The cloning and expression of a cellulase from Bacillus subtilis Y106."
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF355629; AAK39540.1; -
 SO SEQUENCE 499 AA; 55323 MW; 587A687584B7E5BE CRC64;

Query Match 45.2%; Score 364; DB 2; Length 499;
 Best Local Similarity 44.6%; Pred. No. 3.8e-23;
 Matches 66; Conservative 32; Mismatches 46; Indels 4; Gaps 2;

OY 4 GLKVOYKNNDSAPGDNQIKPGLQVNTGSSVDLSVTYVRYWF--TRDGGSTLYVNCDA 62
 DB 353 GISVQYRAGDGSMSNNOIRPOLQIKNNGNTVLDKDYARVYWKAKNKGON--FDCDYA 409
 OY 63 AMCGCNINAFSGSVNPAFTADTYLQLSFTGGTLAAGSGTEIQNRNKSMDMSFETND 122
 DB 410 QIGCGNVTHKFTYLLHKPKQAGDTYLELGFKNGTLPAGASTGNIQLRLHNDMSNVAOSD 469
 OY 123 YSYGNTTFQDMTKVTVVNGVLVWGTE 150
 DB 470 YSEFSKNTFKTKITLYDQGLWTE 497

RESULT 6
 O93JDU0 PRELIMINARY; PRT; 508 AA.
 ID O93JDU0
 AC 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 RT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4).
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CH21;
 RA Zviadzei C.V., Delgado O.D., Zvaunya R., Mattiasson B.;
 RT "Bacillus subtilis CH21 endo-b-1,4-glucanase gene."
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY044252; AAK94871.1; -
 KW Hydrolyase; Glycosidase.
 SO SEQUENCE 508 AA; 56553 MW; A395EELCF340F2 CRC64;

Query Match 45.2%; Score 364; DB 2; Length 508;
 Best Local Similarity 44.6%; Pred. No. 3.9e-23;
 Matches 66; Conservative 32; Mismatches 46; Indels 4; Gaps 2;

OY 4 GLKVOYKNNDSAPGDNQIKPGLQVNTGSSVDLSVTYVRYWF--TRDGGSTLYVNCDA 62
 DB 362 GISVQYRAGDGSMSNNOIRPOLQIKNNGNTVLDKDYARVYWKAKNKGON--FDCDYA 418
 OY 63 AMCGCNINAFSGSVNPAFTADTYLQLSFTGGTLAAGSGTEIQNRNKSMDMSFETND 122
 DB 419 QIGCGNVTHKFTYLLHKPKQAGDTYLELGFKNGTLPAGASTGNIQLRLHNDMSNVAOSD 478
 OY 123 YSYGNTTFQDMTKVTVVNGVLVWGTE 150
 DB 479 YSEFSKNTFKTKITLYDQGLWTE 506

RESULT 7
 O52731 PRELIMINARY; PRT; 499 AA.
 ID O52731
 AC 052731;
 DT 01-JUN-1998 (TREMblrel. 06, Created)
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
 RT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE ENDO-B-1,4-GLUCANASE (EC 3.2.1.4).
 GN CELS.
 OS Bacillus sp. 79-23.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=72363;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-79-23;
 RA Jung K.H., Chun Y.C., Lee J.-C., Kim J.H., Yoon K.-H.;
 RT "Cloning and expression of a Bacillus sp. 79-23 cellulase gene."
 RL Biotechnol. Lett. 18:1077-1082(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-79-23;
 RA Yoon K.-H., Jung K.H.;
 RT Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 DL EMBL: AF045482; AAC02536.1; -
 DR HSP: O85465; 1A3H.
 DR InterPro: IPR001956; Glyco_hydro_F5.
 DR InterPro: IPR001547; Glyco_hydro_F5.
 DR Pfam: PF00942; CBD_3; 1.
 DR ProDom: PD001947; CBD_3; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 KW Hydrolyase; Glycosidase.

DR InterPro: IPR001701; Glyco_hydro_9.
DR Pfam: PF00942; CBD_3; 3.
DR Pfam: PF02011; Glyco_hydro_48; 1.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR PRINTS: PR00844; GLHYDRLASE48.
DR ProDom: PD001947; CBD_3; 2.
DR PROSITE: PD011903; Glyco_hydro_48; 1.
DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
FT NON_TER
SQ SEQUENCE 1711 AA; 189979 MW; E3E987CEB9CD0C21 CRC64;

Query Match 43.2%; Score 348.5; DB 2; Length 1711;
Best Local Similarity 46.1%; Pred. No. 3.5e-21;
Matches 72; Conservative 28; Mismatches 49; Indels 5; Gaps 4;

QY 1 VSGG-LKVOYKNDAPGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLYVNC 59
DB 683 VAGGQIKVLYANKETNSTTIRPMLKVVNTGSSSIDLSRVTIRYWTVDGDKAQSATS- 741
QY 60 DNAMCCGNIIRASFGSVNPRATPTADTYLQLSFTGCT--LAAGSGTEIGIQRKKSQMSNF 117
DB 742 DNAMQIGASVNTFEKFKLVSSVSGADYLLIEGFSGAGQLQGRDGTGEIQIRFKKSQMSNY 801
QY 118 TETNDYSY-GTNTTFODMTKTVYVNGVLWGTE 150
DB 802 NQGNDSWMSQMTNYGENKATVAYIDGVLWGQE 835

RESULT 11
ID 024820 PRELIMINARY; PRT; 1000 AA.
AC 024820;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BETA-GLUCANASE.
OS thermophilic anaerobe M10.
OC Bacteria.
NCBI_Taxid=67756;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M10;
RA Miyake K., Hattori K., Iijima S.;
RT "Characterization of a multi-domain cellulase from an extremely
thermophilic anaerobe strain M10."
RT Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB008029; BAA22939.1; -
DR HSSP: Q06851; INBC.
DR InterPro: IPR001589; Actinln_act_bind.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001000; Glyco_hydro_10.
DR InterPro: IPR001547; Glyco_hydro_F9.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF00331; Glyco_hydro_10; 1.
DR PRINTS: PD00134; GLHYDRLASE10.
DR PRODOM: PD001947; CBD_3; 1.
DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1.
DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
SQ SEQUENCE 1000 AA; 113265 MW; B9F659A56A752C6B CRC64;

Query Match 42.8%; Score 345; DB 2; Length 1000;
Best Local Similarity 46.1%; Pred. No. 3.5e-21;
Matches 70; Conservative 28; Mismatches 50; Indels 4; Gaps 3;
QY 2 SGLKVOYKNDAPGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLYVNCMW 61
DB 372 SGQIKVLYANKETNSTTIRPMLKVVNTGSSSIDLSRVTIRYWTVDGDKAQSATS-DW 430

QY 62 AAMCGNIRASFGSVNPRATPTADTYLQLSFTGCT--LAAGSGTEIGIQRKKSQMSNFTE 119
DB 431 AAGASVNTFEKFKLVSSVSGADYLLIEGFSGAGQLQPKADTGEIQIRFKKSQMSNYNQ 490
QY 120 TNDYSY-GTNTTFODMTKTVYVNGVLWGTE 150
DB 491 GNDMSWISQMTNYGENKATVAYIDGVLWGQE 522

RESULT 12
ID 09X3P5 PRELIMINARY; PRT; 1770 AA.
AC 09X3P5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE XINA.
GN XINA.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoaerobacter group; Caldicellulosiruptor.
NCBI_Taxid=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; Pubmed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
thermophile Caldicellulosiruptor isolate Tok7B.1."
RT Curr. Microbiol. 40:333-340(2000).
DR EMBL: AF078737; AAD30363.1; -
DR HSSP: Q06851; INBC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR003305; CBD_6.
DR InterPro: IPR001000; Glyco_hydro_10.
DR InterPro: IPR002965; P_Rich_extensn.
DR Pfam: PF00942; CBD_3; 3.
DR Pfam: PF02018; CBD_6; 2.
DR Pfam: PF00331; Glyco_hydro_10; 1.
DR PRINTS: PR00134; GLHYDRLASE10.
DR PRINTS: PR01217; PRICHEXTENSN.
DR PRODOM: PD001947; CBD_3; 3.
DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.
SQ SEQUENCE 1770 AA; 193641 MW; 8BAF193704926C92 CRC64;

Query Match 42.6%; Score 343; DB 2; Length 1770;
Best Local Similarity 45.3%; Pred. No. 9.9e-21;
Matches 68; Conservative 29; Mismatches 49; Indels 4; Gaps 3;

QY 4 GLAVQYKNDAPGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLYVNCMAA 63
DB 1104 GLAVLYKNNETSTASTSIRFWFKVYVNGSSVDLSRKIRYWTVDGDKPOSAY-CMAAQ 1162
QY 64 MCGNIRASFGSVNPRATPTADTYLQLSFTGCT--LAAGSGTEIGIQRKKSQMSNFTE 121
DB 1163 IGASVNTFEKFKLVSSVSGADYLLIEGFSGAGQLQPKADTGEIQIRFKKSQMSNYNQAD 1222
QY 122 DYSY-GTNTTFODMTKTVYVNGVLWGTE 150
DB 1223 DMSWISQMTNYGENAKATVAYIDGVLWGQE 1252

RESULT 13
ID 052374 PRELIMINARY; PRT; 1779 AA.
AC 052374;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE FAMILY 10 XLANASE (EC 3.2.1.8).

GN XYNC.
OS Caldicellulosiruptor sp. Rt69B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=70295;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=RT69B.1;
RA Morris D.D., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RT "Family 10 and 11 xylanase genes from Caldicellulosiruptor sp.
RT Rt69B.1.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF036924; AAB95326.1; -
DR HSSP: Q06851; INBC
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR003305; CBD_6.
DR InterPro: IPR001000; Glyco_hydro_10.
DR Pfam: PF00942; CBD_3; 3.
DR Pfam: PF02018; CBD_6; 2.
DR Pfam: PF00331; Glyco_hydro_10; 1.
DR PRINTS: PR00134; GLYDRLASE10.
DR PRODOM: PD001947; CBD_3; 3.
DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.
KW Xylan degradation; Hydrolase; Glycosidase.
SQ SEQUENCE 1779 AA; 194304 MW; CE5269B6806B5CED CRC64;

Query Match 42.3%; Score 341; DB 2; Length 1779;
Best Local Similarity 45.3%; Pred. No. 1.5e-20;
Matches 68; Conservative 27; Mismatches 51; Indels 4; Gaps 3;

QY 4 GLKGYKNNDSAPGDNQIRPGQLVNTGSSSYDLSTVYRVYFTRDGGSSSTLYVNCMDAA 63
DB 1113 GLKVLKNNETASAGSIRPWEKIVNGSSSYDLSTVYRVYFTRDGGSSSTLYVNCMDAA 1171
QY 64 MCGGNIRASFGSVNPATPADIYLOLSTFGT--LAAGSGTEIONRVKSDMSNFTETN 121
DB 1172 IGASVNTFNFVKLSGVSADYLLVGVSSGAGQLQPKGDTDQIVRFKKNDSNYNQAD 1231
QY 122 DYSY-GTNTTFQDMTKVTVYVGVVWGTGTE 150
DB 1232 DMSWQSMNTYGENAKVTLLYVDGVVWGOE 1261

RESULT 14
Q9A0H0 PRELIMINARY; PRT; 996 AA.
AC Q9A0H0;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GLYCOSYL HYDROLASE 5 (FRAGMENT).
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RA MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR HSSP: AF078038; AAK06388.1; -
DR HSSP: Q06851; INBC
DR InterPro: IPR002860; BNR.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR002965; P_Rich_extensn.
DR Pfam: PF02012; BNR; 9.
DR PRINTS: PR01217; PRICHEXTENS.
DR PRODOM: PD001947; CBD_3; 1.

KW Hydrolase.
FT NON_TER 996
SQ SEQUENCE 996 AA; 108275 MW; 3C72B6ED22F3C614 CRC64;

Query Match 42.3%; Score 340; DB 2; Length 996;
Best Local Similarity 45.3%; Pred. No. 9.2e-21;
Matches 68; Conservative 28; Mismatches 50; Indels 4; Gaps 3;

QY 4 GLKGYKNNDSAPGDNQIRPGQLVNTGSSSYDLSTVYRVYFTRDGGSSSTLYVNCMDAA 63
DB 844 GLKVLKNNETASAGSIRPWEKIVNGSSSYDLSTVYRVYFTRDGGSSSTLYVNCMDAA 902
QY 64 MCGGNIRASFGSVNPATPADIYLOLSTFGT--LAAGSGTEIONRVKSDMSNFTETN 121
DB 903 IGASVNTFNFVKLSGVSADYLLVGVSSGAGQLQPKGDTDQIVRFKKNDSNYNQAD 962
QY 122 DYSY-GTNTTFQDMTKVTVYVGVVWGTGTE 150
DB 963 DMSWQSMNTYGENAKVTLLYVDGVVWGOE 992

RESULT 15
Q9A0G4 PRELIMINARY; PRT; 1751 AA.
AC Q9A0G4;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELE.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RA MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL: AF078042; AAK06394.1; -
DR HSSP: Q06851; INBC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001701; Glyco_hydro_9.
DR InterPro: IPR000566; Lipocin_cytfab.
DR InterPro: IPR002965; P_Rich_extensn.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR PRODOM: PD001947; CBD_3; 3.
DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN.1.
SQ SEQUENCE 1751 AA; 192176 MW; 60178CBF3C00B95 CRC64;

Query Match 42.1%; Score 339; DB 2; Length 1751;
Best Local Similarity 44.7%; Pred. No. 2.1e-20;
Matches 67; Conservative 29; Mismatches 50; Indels 4; Gaps 3;

QY 4 GLKGYKNNDSAPGDNQIRPGQLVNTGSSSYDLSTVYRVYFTRDGGSSSTLYVNCMDAA 63
DB 678 GKVLKNNETASAGSIRPWEKIVNGSSSYDLSTVYRVYFTRDGGSSSTLYVNCMDAA 736
QY 64 MCGGNIRASFGSVNPATPADIYLOLSTFGT--LAAGSGTEIONRVKSDMSNFTETN 121
DB 737 IGASVNTFNFVKLSGVSADYLLVGVSSGAGQLQPKGDTDQIVRFKKNDSNYNQAD 796
QY 122 DYSY-GTNTTFQDMTKVTVYVGVVWGTGTE 150

Db 797 DMSHLOSMTNYGENAKVTLTYDGYLVNGQE 826

Search completed: August 29, 2002, 16:37:47
Job time: 957 sec

This Page Blank (uspto)

Tue Sep 3 13:42:03 2002

us-09-917-384-5.ra1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:22:34 ; Search time 47.58 Seconds
(without alignments)
77.004 Million cell updates/sec

Title: US-09-917-384-5
Perfect score: 806
Sequence: 1 VSGGLKVOYKKNNDSPGDNO.....FQDWTKTYVYVGVWGTG 150

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents-AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTCUTS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	367	45.5	551 2 US-09-033-537A-1	Sequence 1, Appl
2	361.5	44.9	700 2 US-07-862-588B-2	Sequence 2, Appl
3	359.5	44.6	493 4 US-09-198-956-10	Sequence 10, Appl
4	359.5	44.6	493 4 US-09-198-955A-12	Sequence 12, Appl
5	358.5	44.5	167 5 PCT-US95-13813-9	Sequence 9, Appl
6	340	42.2	161 4 US-09-136-574A-47	Sequence 47, Appl
7	339	42.1	1751 4 US-09-136-574A-44	Sequence 44, Appl
8	336	41.7	1426 4 US-09-136-574A-43	Sequence 43, Appl
9	329	28.4	162 1 US-08-048-164A-2	Sequence 2, Appl
10	329	28.4	162 1 US-08-460-452-2	Sequence 2, Appl
11	329	28.4	162 1 US-08-460-457-2	Sequence 2, Appl
12	329	28.4	162 1 US-08-460-455-2	Sequence 2, Appl
13	329	28.4	162 2 US-08-330-394A-2	Sequence 7, Appl
14	329	28.4	162 2 US-09-006-633-7	Sequence 22, Appl
15	329	28.4	163 4 US-09-006-633-7	Sequence 22, Appl
16	329	28.4	163 4 US-09-277-716-22	Sequence 22, Appl
17	329	28.4	163 4 US-08-330-394A-22	Sequence 22, Appl
18	326	28.0	156 2 US-08-330-394A-22	Sequence 22, Appl
19	326	28.0	531 2 US-07-862-588B-7	Sequence 7, Appl
20	326	16.1	428 3 US-09-118-319-5	Sequence 7, Appl
21	85	10.0	238 3 US-09-286-690-7	Sequence 7, Appl
22	81	9.9	326 4 US-09-286-691-23	Sequence 23, Appl
23	80	9.9	326 4 US-08-687-147-23	Sequence 23, Appl
24	79	9.8	2123 4 US-08-968-685A-10	Sequence 10, Appl
25	78	9.7	221 1 US-08-315-695-16	Sequence 16, Appl
26	78	9.7	237 1 US-08-103-998-4	Sequence 4, Appl
27	78	9.7		

28	78	9.7	434 2 US-09-008-962-1	Sequence 1, Appl
29	78	9.7	434 2 US-08-675-507-1	Sequence 1, Appl
30	78	9.7	434 3 US-09-213-205-1	Sequence 48, Appl
31	78	9.7	2048 4 US-09-268-347-48	Sequence 5, Appl
32	77.5	9.6	892 4 US-09-336-447A-5	Sequence 6, Appl
33	77	9.6	140 1 US-08-476-275-6	Sequence 9, Appl
34	77	9.6	1338 4 US-08-728-470-9	Sequence 10, Appl
35	77	9.6	1338 4 US-08-719-641-9	Sequence 10, Appl
36	77	9.6	1529 4 US-08-728-470-10	Sequence 9, Appl
37	77	9.6	1529 4 US-08-719-641-10	Sequence 10, Appl
38	77	9.6	1599 2 US-08-617-697-9	Sequence 9, Appl
39	77	9.6	1600 2 US-08-617-697-10	Sequence 10, Appl
40	76.5	9.5	211 1 US-07-842-349-2	Sequence 2, Appl
41	76.5	9.5	211 1 US-08-246-686-2	Sequence 2, Appl
42	76	9.4	625 2 US-08-532-547-7	Sequence 7, Appl
43	76	9.4	625 2 US-08-532-547-9	Sequence 7, Appl
44	76	9.4	625 4 US-09-019-809-7	Sequence 9, Appl
45	76	9.4	625 4 US-09-019-809-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION:
; APPLICANT: Onishi, Masahiro
; APPLICANT: Fich, Merete
; APPLICANT: Toft, Annette Hanne
; APPLICANT: Sh lein, Martin
; TITLE OF INVENTION: Prevention Of Back-Staining
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/033,537A
; FILING DATE: 02-MAR-1998
; CLASSIFICATION: 008
; PRIOR APPLICATION DATA: 0993/95
; APPLICATION NUMBER: 0993/95
; FILING DATE: 08-SEP-1995
; APPLICATION NUMBER: PCT/DK96/00364
; FILING DATE: 03-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Reza
; REGISTRATION NUMBER: 38,475
; REFERENCE/DOCKET NUMBER: 4492,204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 551 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-033-537A-1

Query Match 45.5%; Score 367; DB 2; Length 551;
Best Local Similarity 47.0%; Pred. No. 2,1e-27;
Matches 71; Conservative 31; Mismatches 43; Indels 6; Gaps 4;

QY 2 SGGLYOKKNDNAPGDQIKPGLQVNTGSSVDLSTVYVWFTRDGSSTLYVNC 61
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 402 TGNLVYQKVGSTSTNDQMRKSNFKKNGTTPVNLGSLKRLYTTKD-GRDMSASFPW 460
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

QY 62 AAMGCGNIRASFGSVNPAITPATDYIQLST--GGTLAAGSTGEIQRNKSMSNTE 119
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 461 AQIGSANSVAE--ANFTGSNDTYVELSFGSSGSIPIAGCGTGDILQRLMYKTDMSNFNE 518
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

QY 120 TNDYSY-GTNTFODMTKVTYVYVNGVLWGT 149
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 519 ANDISYDGAKTAYADMNRYTLHNGTLYWGT 549
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 2
US-07-862-588B-2
; Sequence 2, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linna
; APPLICANT: Sch Ieln, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T./ Lambirth, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425, 204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-862-588B-2

Query Match 44.9%; Score 361.5; DB 2; Length 700;
Best Local Similarity 49.4%; Pred. No. 9.6e-27;
Matches 76; Conservative 29; Mismatches 42; Indels 7; Gaps 5;

QY 1 VSGGLKVOYKND-SAPGDQIKPGLQVNTGSSVDLSTVYVWFTRDGSSTLYVNC 59
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 549 VNSDLVVOYKDDRNATQNIKPHNINQKGTSPYDLSLTLRLYFTKD-SSAANNGI 607
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

QY 60 DWAMCGGIRASFGSVNPAITPATDYIQLST--GTLAAGSTGEIQRNKSMSNF 117
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 608 DWAKGGSNIQISFGNHNA--DSDTYAELFGSSGASISGSGEIQLRMSKADMSNF 665
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

QY 118 TETNDYSY-GTNTFODMTKVTYVYVNGVLWGT 150
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 666 NEANDISFGAKTAYIADMRYTLIDQGLWGCIE 699
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 3
US-09-198-956-10
; Sequence 10, Application US/09198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulten, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE REFERENCE: 5377,200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; EARLIER FILING DATE: 1998-11-24
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
; US-09-198-956-10

Query Match 44.6%; Score 359.5; DB 4; Length 493;
Best Local Similarity 44.0%; Pred. No. 9.5e-27;
Matches 70; Conservative 34; Mismatches 44; Indels 11; Gaps 3;

QY 1 VSGGLKVOYKNDNAPGDQIKPGLQVNTGSSVDLSTVYVWFTRDGSSTLYVNC 60
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 331 VSGNLKVEFYNSPDSITNSINPQKRVNTGSSAIDSLTLRYVYVYDQKQDTFW-CD 389
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

QY 61 WAM-----GCGNIRASFGSVNPAITPATDYIQLSTGTLAAGSTGEIQRN 111
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

DB 390 HAAIGNSNGSYGITSNVGTFVKKSSSTNNADTYLEISFTGTLPECAHV-QIOGRFAK 448
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

QY 112 SDMSNTEFTNDYSYGTNTFODMTKVTYVYVNGVLWGT 150
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 449 NDMSNVTOSNDYSFQRSQFVEMDQVTAYLNGVLWGRE 487
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 4
US-09-198-955A-12
; Sequence 12, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulten, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schmitt, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6187580e1 Pectate Lyases
; FILE REFERENCE: 5378, 200-US
; CURRENT APPLICATION NUMBER: US/09/198,955A
; CURRENT FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97

PRIOR FILING DATE: 1997-11-24
 PRIOR APPLICATION NUMBER: 1344/97
 PRIOR FILING DATE: 1997-11-24
 PRIOR APPLICATION NUMBER: 60/067,249
 PRIOR FILING DATE: 1997-12-02
 PRIOR APPLICATION NUMBER: 60/067,240
 PRIOR FILING DATE: 1997-12-02
 PRIOR APPLICATION NUMBER: 09/073,684
 PRIOR FILING DATE: 1998-05-06
 PRIOR APPLICATION NUMBER: 09/184,217
 PRIOR FILING DATE: 1998-11-02
 NUMBER OF SEQ ID NOS: 32
 SOFTWARE: FASTSEQ for Windows Version 4.0
 SEQ ID NO 12
 LENGTH: 493
 TYPE: PRT
 ORGANISM: Clostridium thermocellum
 GS-09-198-955A-12

	Query Match	Best Local Similarity	Match	Conservative	Score 359.5;	DB 4;	Length 493;
QY	1	VSGGLKVOYKKNDDAPGDNIKPGLOLVNTGSSSVLYSVYRWYFTPDGSGSTLYVNC	44.68;	Pred. No. 9.5e-27;	Indels 11;	Gaps 3;	
DB	331	VSGCNKVEFYFNSNPDSDTNINPQFKYTNNGSSAIDLKTLRYYYVDGOKDQTFW-CD	44.08;	Mismatches 44;			
QY	61	MAAM-----GCCNIRASFSGSVNPATPDTADYIOLSTGTTLAAGSTGEIONRYNK					
DB	390	HAALITGNSGYSNGITSNKKGTFFKSSSTNNAADYILEISFTGTLPEGAHV-QIOGRFAK					
QY	112	SDMSNFTETNDYISYGNTEFDQMTKIVYVNGVLWGTG					
DB	449	NDMSNYTQSDMYFSSKRSQFVEMDQVTAIINGVLWGKE					

RESULT 5
 PCT-US95-13813-9
 Sequence 9, Application PC/TUS9513813
 GENERAL INFORMATION:
 APPLICANT: Yeda Research and Development Co. Ltd.
 APPLICANT: Ramot University Authority for Applied
 APPLICANT: Research and Industrial Development Ltd
 APPLICANT: Technion Research and Development Foundation Ltd
 APPLICANT: Bayer, Edward A.
 APPLICANT: Morag, Ely
 APPLICANT: Milchev, Melir
 APPLICANT: Lamed, Raphael
 APPLICANT: Shoham, Yuval
 TITLE OF INVENTION: MODIFIED CELLULOSE-BINDING DOMAIN (CBD)
 TITLE OF INVENTION: PROTEINS AND USE THEREOF
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Brodsky and Nelmark
 STREET: 419 Seventh Street N.W., Ste. 300
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20004
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/13813
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Brodsky, Roger L.
 REGISTRATION NUMBER: 25,618
 REFERENCE/DOCKET NUMBER: BAYER-3

TELEX: 248633 ID NO: 9:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 167 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US95-13813-9

[illegible]

RESULT 6
US-09-136-574A-47
Sequence 47, Application US/09136574A
Patent No. 6294366
GENERAL INFORMATION:
APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland
Berquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.
TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing
Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION DATA:
CLASSIFICATION: <Unknown>
PRIOR APPLICATION NUMBER: US 08/932,571
APPLICATION DATE: September 19, 1997
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>

;; INFORMATION FOR SEQ ID NO: 47:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 616 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-09-136-574A-47

Query Match 42.2%; Score 340; DB 4; Length 616;
Best Local Similarity 43.8%; Pred. No. 9.5e-25;
Matches 67; Conservative 30; Mismatches 52; Indels 4; Gaps 3;

QY 1 VSGIKVQTKNDAPGDNQIKPGQLVNTGSSVDLSTVTVRYWFTRDGSSSTLYVNC 60
DB 1 MSGGKVVYVKNNETSASGSIIRPFKIVNGSSVDLSRVRIRYWTVDGKQSAV-CD 59
QY 61 WAMGCGNIRASFGSVNPAFTADTYLQLSFTGCT--LAAGSTGEIQNRVKSQMSNFT 118
DB 60 WQIGASNTVFVFKLSSGVSADYLLVGFSSGAGQLQPKDGDIDQVRFNKNDMSNYN 119
QY 119 ETNDYSY-GTNTTFQDWTKVTVYVNGVLWGTE 150
DB 120 QADWSMLQSMNTYGENAKVTLYVDGVLWGQE 152

RESULT 7
US-09-136-574A-44
; Sequence 44, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1751 amino acids

;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-136-574A-44

Query Match 42.1%; Score 339; DB 4; Length 1751;
Best Local Similarity 44.7%; Pred. No. 4.5e-24;
Matches 67; Conservative 29; Mismatches 50; Indels 4; Gaps 3;

QY 4 GLKVKYKNDASAPGDNQIKPGQLVNTGSSVDLSTVTVRYWFTRDGSSSTLYVNC 63
DB 678 GKVVYVKNNETSASGSIIRPFKIVNGSSVDLSRVRIRYWTVDGKQSAV-CDWAQ 736
QY 64 MCGNIRASFGSVNPAFTADTYLQLSFTGCT--LAAGSTGEIQNRVKSQMSNFT 121
DB 737 IGASNTVFVFKLSSGVSADYLLVGFSSGAGQLQPKDGDIDQVRFNKNDMSNYN 796
QY 122 DYSY-GTNTTFQDWTKVTVYVNGVLWGTE 150
DB 797 DWSMLQSMNTYGENAKVTLYVDGVLWGQE 826

RESULT 8
US-09-136-574A-43
; Sequence 43, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1426 amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear


```

RESULT 11
US-08-460-457-2
: Sequence 2, Application US/08460457
: Patent No. 5719044
: GENERAL INFORMATION:
: APPLICANT: Shoseyov, Oded
: APPLICANT: Shislegl, Itai
: APPLICANT: Goldstein, Marc A.
: APPLICANT: Dol, Roy H.
: TITLE OF INVENTION: CELLULOSE BINDING DOMAIN FUSION PROTEINS
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/460,457
: FILING DATE: concurrently herewith
: CLASSIFICATION: 435
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/048,164
: FILING DATE: 14-APR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Mistrock, S. Leslie
: REGISTRATION NUMBER: 18,872
: REFERENCE/DOCKET NUMBER: 7809-008
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864/9741
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 162 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-460-457-2

Query Match 28.4%; Score 229; DB 1; Length 162;
Best Local Similarity 35.8%; Pred. No. 7.8e-15;
Matches 58; Conservative 31; Mismatches 55; Indels 18; Gaps 7;

QY 2 SGGIAVQYKNNDSAPFDNDQIKPGLQLVNTGSSSVSDLSIVYRVKFTFDGGSSTILYKNDW 61
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 TSSMSVEEFNSKNSAQTNSTTPIIKITNTSDSLINDVRYRYRYTIDGQIGQTFW-CDH 61
QY 62 AAMGCN-----IRASF--GSVNPATPTADPTVQLSLFPGGLAAGSGTGE--IQNR 108
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 62 AALLGNSTVVDNITSKNTANVKEETASP-TSTYDTIYVEFGFASG--AAITLKNGQPTITQGR 118
QY 109 VKSDMSNFTETINDYSY-GTNTTFQDWTKTVYVNGVLVWGT 149
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 119 ITKSDMSNTQTFNDYSFSDASSSPFVNVNPKVTGYIGAKVLGT 160

RESULT 12
US-08-460-458-2
: Sequence 2, Application US/08460458
: Patent No. 5738984
: GENERAL INFORMATION:
: APPLICANT: Shoseyov, Oded
: TITLE OF INVENTION: KITS AND METHODS OF DETECTION USING CELLULOSE BINDING DOMAIN
: NUMBER OF SEQUENCES: 21

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,458
FILING DATE: concurrently herewith
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/048,164
FILING DATE: 14-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7809-007
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO. 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-460-458-2

Query Match          28.4%  Score 229;  DB 1;  Length 162;
Best Local Similarity 35.8%  Pred. No. 7.8e-15;
Matches 58;  Conservative 31;  Mismatches 35;  Indels 18;  Gaps

QY  2  SGGILVQYKNNDSAPGDNQIKPGLQLVNVTGSSSVLDSTVYVYWFPTRDGSSITLYNCDM 61
   :  |::|::|  :  |::|::|  :  |::|::|  :  |::|::|  :  |::|::|  :  |::|::|
DB  3  TSMSEVFEFNSKMSQNTSITPIIKITNTSDSDLNDVKVRYTSDGTQGTFTW-CDH 61
   :  |::|::|  :  |::|::|  :  |::|::|  :  |::|::|  :  |::|::|  :  |::|::|
QY  62  AAMGCGN-----IRASF--GSVPATPTADTYIQLSFTGCTLAAGSTGE---IQNR 108
   :  |::|::|  :  |::|::|  :  |::|::|  :  |::|::|  :  |::|::|  :  |::|::|
DB  62  AGALLGNSVVDNTSKVTAFNFKETASP-TSTYDYVEEFGFASG--AATLKKGQIFITQGR 118
   :  |::|::|  :  |::|::|  :  |::|::|  :  |::|::|  :  |::|::|  :  |::|::|
QY  109  VAKSDMSNFTETNDYSY-GINTTFQDWFRTVYVYVNGVLWGT 149
   :  |::|::|  :  |::|::|  :  |::|::|  :  |::|::|  :  |::|::|  :  |::|::|
DB  119  ITKSDMSNTQTNDYSFDASSSTPYVNFKVTGIGGAKYLVGT 160
   :  |::|::|  :  |::|::|  :  |::|::|  :  |::|::|  :  |::|::|  :  |::|::|

RESULT 13
US-08-460-455-2
: Sequence 2, Application US/08460455
: Patent No. 5837814
:
: GENERAL INFORMATION:
: APPLICANT: Shoseyov, Oded
: APPLICANT: Shpilegl, Itai
: APPLICANT: Goldstein, Marc A.
: APPLICANT: Dol, Roy H.
: TITLE OF INVENTION: CELLULOSE BINDING DOMAIN PROTEINS
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,455
FILING DATE: concurrently herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/048,164
FILING DATE: 14-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7809-009
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 869-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-460-455-2

Query Match 28.4%; Score 229; DB 2; Length 162;
Best Local Similarity 35.8%; Pred. No. 7.8e-15;

Matches 58; Conservative 31; Mismatches 55; Indels 18; Gaps 7;

QY 2 SGGKAYKNNDSAPGDNQIKPGLQVNTGSSVDLSTVYRVYFTFRDGGSTLYVNCW 61
DB 3 TSSMSVEFYNSKNSAQTNSTPIIKITNTSDSLNDVRYVYTSDBGQGTFFW-CDH 61
QY 62 AAMGCN-----IRASF--GSVNPATPTADTYLQLSFTGTLAAGSGTGE---IQNR 108
DB 62 AGALLGNSYVDNISKYANVKETASP-TSTYDYVEFGFASG--AATLKKGQFTITQGR 118
QY 109 VNKSDFSNETNDYSY-GTNTTFQDMTKYVYVNGVLVWGT 149
DB 119 ITRSDMSNTQTNDYSFSDASSSTPVVNPVKYTGIGAKVLTG 160

RESULT 14

US-08-330-394A-2

Sequence 2, Application US/08330394A

Patent No. 5856201

GENERAL INFORMATION:

APPLICANT: Shoseyov, Oded

APPLICANT: Yosef, Karmey

APPLICANT: Shpiegel, Itai

APPLICANT: Goldstein, Marc A.

APPLICANT: Dol, Roy H.

TITLE OF INVENTION: METHODS OF DETECTION USING THE

TITLE OF INVENTION: CELLULOSE BINDING DOMAIN

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS

STREET: 1155 AVENUE OF THE AMERICAS

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/330,394A

FILING DATE: 27-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: CIP OF PCT/US94/04132
FILING DATE: 14-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: MISROCK, S. LESLIE
REFERENCE/DOCKET NUMBER: 7809-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66441 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-330-394A-2

Query Match 28.4%; Score 229; DB 2; Length 162;

Best Local Similarity 35.8%; Pred. No. 7.8e-15;

Matches 58; Conservative 31; Mismatches 55; Indels 18; Gaps 7;

QY 2 SGGKAYKNNDSAPGDNQIKPGLQVNTGSSVDLSTVYRVYFTFRDGGSTLYVNCW 61
DB 3 TSSMSVEFYNSKNSAQTNSTPIIKITNTSDSLNDVRYVYTSDBGQGTFFW-CDH 61
QY 62 AAMGCN-----IRASF--GSVNPATPTADTYLQLSFTGTLAAGSGTGE---IQNR 108
DB 62 AGALLGNSYVDNISKYANVKETASP-TSTYDYVEFGFASG--AATLKKGQFTITQGR 118
QY 109 VNKSDFSNETNDYSY-GTNTTFQDMTKYVYVNGVLVWGT 149
DB 119 ITRSDMSNTQTNDYSFSDASSSTPVVNPVKYTGIGAKVLTG 160

RESULT 15

US-09-006-636-7

Sequence 7, Application US/09006636

Patent No. 6005092

GENERAL INFORMATION:

APPLICANT: Shoseyov, Oded

APPLICANT: Shani, Ziv

TITLE OF INVENTION: ARABIDOPSIS THALIANA ENDO-1,4-BETA-

TITLE OF INVENTION: GLUCANASE GENE, PROMOTER AND PROTEIN

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/006,636

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Baldwin, Geraldine F.

REGISTRATION NUMBER: 31,232

REFERENCE/DOCKET NUMBER: 7809-019

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 163 amino acids

TYPE: amino acid

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:21:43 ; Search time 127 Seconds
(without alignments)
912.204 Million cell updates/sec

Title: US-09-917-384-6
Sequence: 1 MERTQOSGRNCRVGRGTRM.....GFNGSYSGTNAPTLSCRAS 1043

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	1023	18.5	1010	19 AAM34989
2	1000	18.1	596	17 AAR90715
3	999	18.0	423	17 AAR90716
4	646.5	11.7	381	20 AAY06368
5	646.5	11.7	381	21 AAB14881
6	637	11.5	386	21 AAY67497
7	635.5	11.5	371	20 AAY06367
8	635.5	11.5	371	21 AAB14880
9	635.5	11.5	371	21 AAY84345
10	635	11.5	429	21 AAY84346
11	633.5	11.4	371	20 AAY08473

12	633.5	11.4	371	21 AAY67496
13	574.5	10.4	395	21 AAY84797
14	573	10.4	473	19 AAY01077
15	565.5	10.2	476	19 AAM44853
16	565.5	10.2	476	19 AAM44827
17	548	9.9	471	6 AAY50308
18	540	9.8	471	16 AAR77262
19	540	9.8	471	17 AAM02025
20	539	9.7	457	18 AAM25789
21	472.5	8.5	446	20 AAY01076
22	467	8.4	1751	20 AAY13493
23	449	8.1	1426	20 AAY13492
24	444.5	8.0	449	19 AAM56739
25	431.5	7.8	432	21 AAB18414
26	417.5	7.5	551	18 AAB18750
27	414	7.5	562	21 AAY69508
28	413	7.5	1853	19 AAM43108
29	412	7.4	459	19 AAM56738
30	410.5	7.4	521	22 AAB48788
31	407.5	7.4	260	20 AAY06369
32	407.5	7.4	260	21 AAB14882
33	407.5	7.4	260	21 AAY84347
34	406	7.3	521	17 AAB89927
35	406	7.3	521	22 AAB48786
36	405.5	7.3	782	12 AAB15625
37	404.5	7.3	476	21 AAY54123
38	393	7.1	700	12 AAR13227
39	388.5	6.9	499	14 AAR42122
40	383	6.9	431	18 AAB15238
41	381.5	6.9	531	18 AAG63962
42	377	6.8	1352	22 AAG63962
43	376	6.8	616	20 AAY13494
44	373.5	6.7	1350	22 AAG63963
45	371.5	6.7	493	20 AAY28850

ALIGNMENTS

RESULT 1
AAM34989 standard; Protein; 1010 AA.
AAM34989:
AC AAM34989:
DT 21-MAY-1998 (first entry)
XX
XX
XX Tereidnibacter endoglucanase.
XX Endoglucanase: cellulase; carboxymethylcellulose; cellulose;
XX Biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;
XX thermotable enzyme; thermophilic; glycosidase.
OS Tereidnibacter sp. (Clone 426P1).
XX
XX
XX WO9744361-A1.
PN
XX
XX 27-NOV-1997.
PD
XX
XX 22-MAY-1997; 97WO-US08793.
PE
XX
XX 22-MAY-1996; 96US-0651572.
PR
XX (RECO-) RECOMBINANT BIOCATALYSIS INC.
PA
XX Lam DE, Mathur EJ;
PI
XX WPI: 1998-018435/02.
DR N-PSDB; AAT44197.
DR
XX Endoglucanase(s), preferably form archaeal bacterium, AEP1 1a
PT useful to degrade carboxymethylcellulose and hydrolyse of
PT beta-1,4-glycosidic bonds in cellulose

Actinomyces cell
Amino acid sequenc
H. insolens cell
Humicola insolens
H. insolens cell
Cellulohydrolytic
T. longibrachiatum
Trichoderma cell
Acetomolus cell
H. insolens cell
Truncated cellulas
Orpinomyces cell
Amino acid sequenc
Corrected Bacillus
Acidothermus cell
C. thermocellum ce
Orpinomyces cell
Acidothermus cell
Rhodothermus marit
Emeritella desarto
A. cellulolyticus
Acidothermus cell
Cellulase AE-1. A
Acidothermus cell
Novel endoglucanas
Scaffoldin protein
Amino acid sequenc
Truncated cellulas
Amino acid sequenc
Pectate lyase-link

XX Claim 1; Fig 1E; 16app; English.

PS This protein comprises an endoglucanase of Terebrinibacter (clone

CC 432P1) that is capable of degrading carboxymethylcellulose and of

CC hydrolysing the beta-1,4-glycosidic bonds in cellulose. It has

CC homology to an endoglucanase of archaeobacterium APB11a (see

CC AAM34985). It can be produced from native cells or from recombinant

CC host cells, especially prokaryotic host cells transformed with a

CC plasmid or virus-derived vector including the endoglucanase DNA

CC (see AAT94197). 24 Endoglucanases (see AAM34986-W35008) are claimed.

CC They can be used to degrade cellulose for the conversion of plant

CC biomass into fuels and chemicals, for use in detergents, textiles,

CC animal feed, waste treatment, and in the fruit juice and brewing

CC industries for the clarification and extraction of juices.

XX Sequence 1010 AA;

Query Match 18.5%; Score 1023; DB 19; Length 1010;

Best Local Similarity 48.5%; Pred. No. 7.3e-57;

Matches 210; Conservative 68; Mismatches 121; Indels 34; Gaps 13;

57 VDNPPAGA-TFEVNPYMAOEVOSEANOTNATLAKMRVYSTYAVWMDRTAINGVNG 115

DB 598 vdnppagaqkwyinpmws-----asaaneppgs-----vianeppfwmndrigalepad 647

OY 116 GPGLTFTYIDALISQOOGTPEVIEIYIDLPGRDCAALASNGELPATAGLQTYETQYID 175

DB 648 gmgldrdhnealag----gadlfmfvvydlpurdcaalangelrlsedgfnlyksdyla 703

OY 176 PLASTLSNPKYSLSLIVYIIIEPDSLPMATNMSIQAC--ATNVPYEGGIEFALTKLNI 233

DB 704 plvelslpdpayaglklaavleavslpnlvtlnsepdceangpggyrdgirnatefgk 763

OY 234 PNVTIYMDAHSGLGMPNN-ASGYVOEYOKYLNASIGVNGIDGVTNANTPTKEPPM 292

DB 764 pnyvtyvdalshgwlgnndfnagynlllyevanlgsnglnpdlagfysnasytprpepf 823

OY 233 -TATQOOGQPYESANFYQWNPDIADYAVDLVSRVLAAGFPSSIGMLITLNGWGP 351

DB 824 pdanlqvgqpyrvasdfyewnsylaekepfvltwrsamlskmpsslgmldlarngwgp 883

OY 352 NEPTGPSTATDVNTFVNOSKIDLRHGLMCNONGAGLGPQASPTDEPNMHLAYVI 411

DB 884 erptagsismnltvnesridrtretrgnwcnpg-gvgyrptaa---pspgldayvvw 938

OY 412 KPPGESDGTSAAS--DPT-TGKKSDDPMCDP---TYTTSYGLTNALPNSPIAGOMFPA 463

DB 939 kpggesdgsdpnfelddpndpdkqndpmcdpfasnsnsayg--tgampnphagrtwfp 996

OY 464 QPDQLVANARPAV 476

DB 997 aqqlllenayppl 1009

RESULT 2

AAR90715

ID AAR90715 standard; Protein; 596 AA.

XX AAR90715;

XX 06-APR-1996 (first entry)

XX Thermostable cellulase-E3.

XX Cellulase-E3; Thermomonospora fusca; thermostable; plasmid pS24;

XX plasmid pS26; Escherichia coli; cloning; Streptomyces lividans;

XX papain; catalytic domain; cellulose; chitosan; protease resistance;

XX synergism; cellobiohydrolase; beta-glucosidase; saccharification;

XX surfactant additive; paper recycling; delinking; paper refining.

OS Thermomonospora fusca strain YX36.

XX Key Location/Qualifiers

FT Peptide 1..38

FT Peptide /note= "Signal peptide"

FT Peptide 39..48

FT Domain /note= "N-terminal peptide (AAR90717)"

FT Domain 174..596

PN WO9600281-A1 /note= "Catalytically active domain (AAR90716)"

XX 04-JAN-1996.

XX 23-JUN-1995; 95WO-US09069.

XX 24-JUN-1994; 94US-0265429.

XX (CORR) CORNELL RES FOUND INC.

XX Walker LP, Wilson DB, Zhang S;

XX WPI; 1996-068865/07.

XX N-PSDB; AAT15595.

XX Cellulase from Thermomonospora fusca - is thermostable and is useful

XX in a variety of industrial applications e.g. clarification of fruit

XX juices, fabric softening, etc

XX Claim 1; Page 32-37; 53pp; English.

XX The sequence represents thermostable cellulase-E3 (EC-3.2.1.4) from

XX Thermomonospora fusca, and is encoded by inserts in plasmid pS24

XX and plasmid pS26, obtained by cloning in Escherichia coli. The

XX gene may be cloned and expressed in E. coli or Streptomyces

XX lividans for recombinant cellulase-E3 production. Cellulase-E3,

XX or a catalytically active polypeptide obtained by papain digestion,

XX hydrolyses cellulose or chitosan at pH 5-11 and 40-70 deg C, and

XX has significant activity at over 60 deg C. E3 has higher stability

XX to proteolysis in culture supernatant than T. fusca cellulases-E2

XX and -E5, and shows strong synergistic activity when combined with

XX other cellulases, cellobiohydrolase and optionally beta-glucosidase.

XX The cellulase and mixtures may be used e.g. in cellulose

XX saccharification for ethanol production, fruit juice clarification,

XX as a surfactant additive, or in delinking or refining of recycled

XX paper.

XX Sequence 596 AA;

Query Match 18.1%; Score 1000; DB 17; Length 596;

Best Local Similarity 47.7%; Pred. No. 1.1e-55;

Matches 210; Conservative 55; Mismatches 143; Indels 32; Gaps 9;

OY 50 PAIAATHVDNPPAGATFEVNPYMAOEVOSEANOTNATLAKMRVYSTYAVWMDRTA 109

DB 172 ptnpekevpndieagklvynpwsakaaepgs-----avaestavldrtiga 221

OY 110 ING---VNGGPGLTFTYIDALISQOOGTPEVIEIYIDLPGRDCAALASNGELPATAG 165

DB 222 legndspitgsmglrdhleav-rqsggdplctlgvlynlpurdcaalangelpdpd-- 278

OY 166 LOTYETQYIDPIASTLSN-PRYSSLRIVYIIIEPDSLPMATNMSIQACATAVPY----- 218

DB 279 ldrkyseyldpladlmwdfadyenlrivallidsjplvtntvgngygtelcaymkqng 338

OY 219 YEGGIEYALTKHAIPNVYIYMDAHSGLGMPNNSAGYVOEYOKYLNAS-IGVNGIOGF 277

DB 339 yngvgyalrkigelpnyvnyidaahgwlwdsnfgpsvdlfyeeanaasgatyvngf 398

OY 278 VNTVNTNPTPLKEPFMTATQOOGQPYESANFYQWNPDIADYAVDLVSRVLAAGFPSSIT 337

DB 399 lantanyasatvepyldvngvtngqlirgskwvdmngyvdeisfvqdlrqaallakfrsdi 458

Oy	338	GLMLDTLRNKGSGNEPTGPESTADVTMTFVNQSKIDLRHRLGMCNONGAGLGQPPQASP	397		
Db	459	gmlldtstrngvggnrncrlcgssscldnlcydesldtrllhpgnmcnqagslgerlvpmp	518		
Oy	398	TDFPNALDAVAVWVKPKPGESDXTGSAADPTTGKKSDMPCDPTWTSTGYGLTN--ALPNS	454		
Db	519	a----pgvdavyywvkprgeesgaaseelpndegkfgldrmcdpptygnarrnnpnpsgalpna	574		
Oy	455	PIAGQWPPAGPDQLVANARP	474		
Db	575	plsghwfsaqfrellanayp	594		
RESULT 3					
ID	AA890716	standard; Protein: 423 AA.			
XX	AA890716:				
DT	06-APR-1996	(first entry)			
XX	Thermostable cellulase-E3 catalytic domain.				
DE					
XX	Cellulase-E3: Thermomonospora fusca; thermostable; papain;				
KW	catalytic domain; Escherichia coli; cloning; Streptomyces lividans;				
KM	cellulose; chitosan; protease resistance; synergism;				
KM	cellobiohydrolase; beta-glucosidase; saccharification;				
KM	surfactant additive; paper recycling; deinking; paper refining.				
XX					
OS	Thermomonospora fusca strain YX36.				
XX					
PN	WC9600281-A1.				
PD	04-JAN-1996.				
XX					
PF	23-JUN-1995; 95MO-US09069.				
XX					
PR	24-JUN-1994; 94US-0265429.				
XX					
PA	(CORR) CORNELL RES FOUND INC.				
PI	Walker LP, Wilson DB, Zhang S;				
XX					
DR	WPI; 1996-068865/07.				
DR	N-PSDB; AAT15596.				
XX					
PT	Cellulase from Thermomonospora fusca - is thermostable and is useful				
PT	in a variety of industrial applications e.g. clarification of fruit				
PT	juices, fabric softening, etc				
XX					
PS	Claim 1; Page 38-40; 53pp; English.				
XX					
CC	The sequence corresponds to a catalytically active domain of				
CC	thermostable cellulase-E3 (EC-3.2.1.4) from Thermomonospora				
CC	fusca, and is isolated by papain cleavage of the full-length				
CC	protein (AA890715). The binding domain and linker region have				
CC	been removed from the full-length protein. The protein may be				
CC	produced in recombinant form in Escherichia coli or Streptomyces				
CC	lividans. The domain retains full activity, but shows reduced				
CC	cellulose binding. The catalytic domain, or the corresponding				
CC	full-length cellulase, hydrolyses cellulose or chitosan at pH				
CC	5-11 and 40-70 deg C, and has significant activity at over 60				
CC	deg C. E3 has higher stability to proteolysis in culture				
CC	supernatant than T. fusca cellulases-E2 and -E5, and shows strong				
CC	synergistic activity when combined with other cellulases,				
CC	cellobiohydrolase and optionally beta-glucosidase. The cellulase				
CC	and mixtures may be used e.g. in cellulose saccharification for				
CC	ethanol production, fruit juice clarification, as a surfactant				
CC	additive, or in deinking or refining of recycled paper.				
XX					
Sequence	423 AA;				
XX					

Query Match	18.0%	Score 999	DB 17	Length 423
Best Local Similarity	48.3%	Pred. No. 7.7e-56		
Matches 209	Conservative 55	Mismatches 137	Indels 32	Gaps 9
OY	57	VDNPRAGATFFFNPRWAQBOVESEANQGNATLAAKMRVSTYSTFVMMDRIAING----	112	
Db	6	vdnprgagklyllynpwsakaaepggs-----	avanesstavwldrlgalegnasp	55
OY	113	VNGSGELTYYTLDAAISOQOGTTPVEYIEIVYDLPGRDCAALASNGELPATAGLOTYETQ	172	
Db	56	ttgsmglrdhleeav-rqsggdrltqylylnlpdrccaalasngelppde--ldrylse	112	
OY	173	YIDPIASTLSN-PKYSLSRLVTIIEPDSLPANAVTKMSIOACATAVPY-----YEOGIEY	225	
Db	113	yldprradlmwdfadyenrlrtvalleldslrlnltvngnggrelcalymkqngyynvgy	172	
OY	226	ALTKHAHPNYYIYMDAAHSGWLGRPNNASGYOEVOKVLNMS-IGVNGIDGFYNTNTAY	284	
Db	173	alktgehpnyyylidaahgwlygdnfgrpsvdlfyeaanaagstvdvnhflsantany	232	
OY	285	TPLEKPEMTATQOVGCGOPVESANFYQMPMPDIDEADYAYDLSRLVAAAGPSSIGMLIPTL	344	
Db	233	satvepyldvngtvtngqllrsgskwdmngyvdelstfvqdlrtaallakgrsfslgnldts	292	
OY	345	RNGWGNPEPTGPRSTATDVNTFFVNOSKIDLRORHRLMCONCAGAGLORPOASPTDFPNAH	404	
Db	293	rngwgnprtrgpgssstldnlyvdesrldrtlrhpgnwcwagajgserptvnpa----pg	348	
OY	405	LDAYWIRPFGSGDSGTSAASDPPTTKKSDPMCKDPRYYTTSYGVLLN---ALNPSPLAGOWF	461	
Db	349	vdaywvwrkpgesdgaseelpndegkfgtdfcmcdprygnatrgnnpagallnaprlsghwf	408	
OY	462	PAQFDQLVANARP	474	
Db	409	saftrellanayp	421	

RESULT	4
AA06368	
ID	AA06368 standard; Protein; 381 AA.
XX	
AC	AA06368;
XX	
DT	06-SEP-1999 (first entry)
XX	
DE	Streptomyces lividans EGIII-like cellulase.
XX	
KW	Cellulase; endoglucanase; EGIII; textile; feed additive; baking;
KW	food processing; grain wet milling; pulp; paper.
XX	
OS	Streptomyces lividans.
XX	
PN	W09931255-A2.
XX	
PD	24-JUN-1999.
XX	
PE	14-DEC-1998; 98WO-US26552.
XX	
PR	16-DEC-1997; 97US-0991720.
XX	
PA	(GENV) GENENCOR INT INC.
XX	
PI	Bower BS, Fowler T, Phillips JI;
XX	
DR	WPI; 1999-395187/33.
XX	
PT	EGIII like cellulase
XX	
PS	Example; Fig 6; 47pp; English.

The present polypeptide represents a full-length sequence of a novel EGIII-like cellulase of *Streptomyces lividans* (CeIB). It was deduced from a gene sequence isolated from genomic DNA using PCR

CC primers (see AAY59180-91) based on conserved motifs (see AAY06325-29)
CC of Trichoderma reesei EGI11 cellulase and related enzymes. PCR
CC has been used to identify novel EGI11-like enzymes, including the
CC present protein, from bacterial and fungal sources (see AAY06331-70).
CC The sequence shows homology to T. reesei EGI11 (see AAY06330). Also
CC provided by the invention are vectors, host cells and methods
CC for the recombinant production of such enzymes, which can be used
CC in the treatment of cellulose-containing textiles, as feed of
CC additives, in the treatment of wood pulp, in the reduction of
CC biomass to glucose, in the stone washing of indigo dyed denim, or
CC as laundry detergent components (all claimed).
SQ Sequence 381 AA:

Query Match 11.7%; Score 646.5; DB 20; Length 381;
Best Local Similarity 40.7%; Pred. No. 2.4e-33;
Matches 155; Conservative 45; Mismatches 144; Indels 37; Gaps 14;

QY 684 GLAAG-----TAYTYVAAVDAAGNTSAPSTPVDCTPGPNONGVYSVQDGEYRVQTNEMN 738
DB 13 gllaa1gavlaafalvsslvtaapagadt--lcep---fgtltldg-gryvvgnnrwg 65

QY 739 SSAOQCLTINTATGAMTVSTANFSGGTGAPATPSPISYKCHMGNCTTKNVGMPIDISQI 798
DB 66 stapqcvta-ldtg-frvqadgssapltngapkyspvsfngchync-spgtldpvrldtv 122

QY 799 GSAYTSMSTTOYSSGAYDVAIDWNTNSPTTGTGPNGTIMIMLSRGGVOPFGSQTARG 858
DB 123 saapsislygfvdgavynasydlwldptarldg-vnqtelimwfnrvspgldqigsrvgt- 180

QY 859 VYVAGHTMVMVMOGQQTSMKIIISVLTLPGATISNLDLKAFIDMAARGSLNSTDYLDVE 918
DB 181 asvgrtfrtewwsggnsndvisfv-apsalsgwsfwmfdvratvarglaendwyltsvq 239

QY 919 AGEFIWOGGQGLGNSFSVSV---TSGGV-----ACRATYVYVNSDMGSGFTATTV 966
DB 240 agfepwqngaglavnsfsstveltpgtdpdpdgpsacavsytnv-wqdgftadv 298

QY 967 TNTGSRATNGMTVAMSGFGNGOTVNTWNTALTQSGASVTATNLISNNVIOPOGSTTGFN 1026
DB 299 tntglapvqgwqlatflpsqgrltanwnasltpsgsyvatgatgashariapgsisfgtg 358

QY 1027 GSYSGTNAAPT-----LSTCT 1041
DB 359 gcyggaafaepctgfrlntact 379

RESULT 5
AAB14881
ID AAB14881 standard; Protein; 381 AA.
XX
AC AAB14881;
XX
DT 21-NOV-2000 (first entry)
XX
XX Rhodothermus marinus EGI11-like cellulase.
DE
XX Rhodothermus marinus; Trichoderma reesei; endoglucanase III; EGI11;
KW cellulase; mutant; enzyme stability; textile treatment;
KM wood pulp treatment; feed additive; detergent.
XX
OS Rhodothermus marinus.
XX
PN WO200037614-A2.
XX
PD 29-JUN-2000.
XX
PF 12-NOV-1999; 99WO-US26704.
XX
PR 18-DEC-1998; 98US-0216295.
XX
PA (GENV) GENENCOR INT INC.

XX
PI Mitchinson C, Wendt DJ;
XX
DR WPI, 2000-482483/42.
XX
XX
PT Novel endoglucanase III or endoglucanase III-like cellulase useful for
PT treating textiles and wood pulp comprises a substitution or deletion at
PT specified positions in the wild form of endoglucanase III -
XX
XX
PS Example 1; Fig 3; 52pp; English.
XX
CC The present sequence is a cellulase related to endoglucanase III (EGI11)
CC from Trichoderma reesei. EGI11-like genes were isolated from genomic DNA
CC libraries constructed from various microorganisms by PCR. The isolated
CC genes showed significant homology to EGI11 from T. reesei. Certain
CC substitution and deletion mutations have been incorporated into EGI11 and
CC EGI11-like cellulases to produce variant enzymes with improved stability,
CC e.g. increased resistance to temperature stress. The mutants may be used
CC in textile and wood pulp treatment, as a feed additive, and for reducing
CC biomass to glucose. They are also useful for stonewashing or indigo dyed
CC denim and as an agent in laundry and dish detergents.
SQ Sequence 381 AA:

Query Match 11.7%; Score 646.5; DB 21; Length 381;
Best Local Similarity 40.7%; Pred. No. 2.4e-33;
Matches 155; Conservative 45; Mismatches 144; Indels 37; Gaps 14;

QY 684 GLAAG-----TAYTYVAAVDAAGNTSAPSTPVDCTPGPNONGVYSVQDGEYRVQTNEMN 738
DB 13 gllaa1gavlaafalvsslvtaapagadt--lcep---fgtltldg-gryvvgnnrwg 65

QY 739 SSAOQCLTINTATGAMTVSTANFSGGTGAPATPSPISYKCHMGNCTTKNVGMPIDISQI 798
DB 66 stapqcvta-ldtg-frvqadgssapltngapkyspvsfngchync-spgtldpvrldtv 122

QY 799 GSAYTSMSTTOYSSGAYDVAIDWNTNSPTTGTGPNGTIMIMLSRGGVOPFGSQTARG 858
DB 123 saapsislygfvdgavynasydlwldptarldg-vnqtelimwfnrvspgldqigsrvgt- 180

QY 859 VYVAGHTMVMVMOGQQTSMKIIISVLTLPGATISNLDLKAFIDMAARGSLNSTDYLDVE 918
DB 181 asvgrtfrtewwsggnsndvisfv-apsalsgwsfwmfdvratvarglaendwyltsvq 239

QY 919 AGEFIWOGGQGLGNSFSVSV---TSGGV-----ACRATYVYVNSDMGSGFTATTV 966
DB 240 agfepwqngaglavnsfsstveltpgtdpdpdgpsacavsytnv-wqdgftadv 298

QY 967 TNTGSRATNGMTVAMSGFGNGOTVNTWNTALTQSGASVTATNLISNNVIOPOGSTTGFN 1026
DB 299 tntglapvqgwqlatflpsqgrltanwnasltpsgsyvatgatgashariapgsisfgtg 358

QY 1027 GSYSGTNAAPT-----LSTCT 1041
DB 359 gcyggaafaepctgfrlntact 379

RESULT 6
AAY67497
ID AAY67497 standard; Protein; 386 AA.
XX
AC AAY67497;
XX
DT 19-MAY-2000 (first entry)
XX
DE Protein sequence of Cella and cellulase 11AG8 fusion.
XX
KW Cellulase; Actinomycetes; detergent; feed additive; textile treatment;
KM pulp; paper; cellulase 11AG8.
XX
OS Streptomyces lividans.
XX
XX Synthetic.

XX Key Location/Qualifiers
 PH Peptide 1..46
 FT /note= "celA signal sequence"
 FT 47..386
 FT Protein /note= "cellulase 11A68 mature peptide"
 XX WO200009707-A1.
 XX 24-FEB-2000.
 XX 28-MAY-1999; 99WO-US11971.
 XX 24-JUN-1998; 98US-0104308.
 XX 18-NOV-1998; 98MO-US24649.
 XX 28-MAY-1999; 99US-0321981.
 XX (GEMV) GENENCOR INT INC.
 XX Jones BE, Van Der Kleij WAH, Van Solingen P, Weyler W;
 XX WPI: 2000-224344/19.
 XX N-PSDB: AAZ57031.
 XX A novel Actinomycetes cellulase and related DNA, useful for detergent
 XX compositions, treating textiles and paper or pulp
 XX Example 6; Fig 15; 72pp; English.
 XX The invention provides a cellulase from Actinomycetes. The cellulase can
 XX be used in a detergent composition, as an additive for animal feed and
 XX for the treatment of textiles or pulp and paper. The DNA encoding the
 XX cellulase can be used to identify homologous cellulases and for
 XX recombinant production of cellulases. The present sequence represents
 XX the protein sequence of a celA signal sequence and cellulase 11A68
 XX the fusion sequence contained in the expression cassette consisting of the
 XX GI promoter, celA signal sequence, cellulase 11A68 and GI terminator
 XX Sequence 386 AA:
 SQ
 Query Match 11.5%; Score 637; DB 21; Length 386;
 Best Local Similarity 40.6%; Pred. No. 9.8e-33;
 Matches 149; Conservative 48; Mismatches 134; Indels 36; Gaps 15;
 QY 698 VDAAGNTSAPSTPYDCTPPGPNQ-----GVTSDGEYRQTNEMNSSAQCCLTINTATG 752
 DB 31 vsivglitalagppaga-----nqgldrygtltlqg-ryvvqnmrwgtsatqg--lnvtgn 83
 QY 753 AMVTSTANFSGGTGAPATYPSIYKGCWMGNCCTTKNVGMPIQISQISAVTSMTQVSS 812
 DB 84 gfeitqadqsvplngapksydvqchgnacprt-1pmrlslsapsvsvyrytgn 142
 QY 813 GAVDAVYDWTNSTPTTGTGPNGTETIMTWLNSRGVOPFGSQRTGTVAGHTWNNWQO 872
 DB 143 gvyuaeydlwldptprtg-vnrteimwfnrvpvpqipsvgt-ahvgsrsewvrg 200
 QY 873 QTSWKIISYVLTFGATISINLDKATFADAARGSLNTSD-YLDDVEAGFEIWOGGGIG 931
 DB 201 ngndvlsif-lapsalswsvfdvkd-fvdqavshglapdwyltalsigafepwegegt 258
 QY 932 SNSFSVSVTSGG-----VACRAITYVNSDMGSGFATYVNTGSRATNGTWVMS 982
 DB 259 vnsfsaavnnagngvtpgtspacqysahlt-wpgtftcltclntgspdweldft 317
 QY 983 FCGNQTYTNWNTALNOSGASVTATNLXSN-NVIOGQSTTFEGNCSYSGTN-AAPT--- 1037
 DB 318 lpaghtlvsawmalispsagvatarstgngriiaangtqsfqgtssqgfnapaggr 377
 QY 1038 ---LSCT 1041
 DB 378 lngtact 384

RESULT 7
 ID AAY06367 standard; Protein; 371 AA.
 AC AAY06367;
 DT 06-SEP-1999 (first entry)
 DE Streptomyces sp. EgIII-like cellulase.
 KW Cellulase; endoglucanase; EgIII; textile; feed additive; baking;
 KW food processing; grain wet milling; pulp; paper.
 OS Streptomyces sp.
 PN WO931255-A2.
 PD 24-JUN-1999.
 PF 14-DEC-1998; 98MO-US26552.
 PR 16-DEC-1997; 97US-0991720.
 PA (GEMV) GENENCOR INT INC.
 PI Bower BS, Fowler T, Phillips JT;
 DR WPI: 1999-395187/33.
 XX EgIII like cellulase
 XX Example: Fig 6; 47pp; English.
 XX The present polypeptide represents a full-length sequence of a
 XX novel EgIII-like cellulase of Streptomyces sp. 11A68. It was
 XX deduced from a gene sequence isolated from genomic DNA using PCR
 XX primers (see AAX59180-91) based on conserved motifs (see AAY06325-29)
 XX of Trichoderma reesei EgIII cellulase and related enzymes. PCR
 XX has been used to identify novel EgIII-like enzymes, including the
 XX present protein, from bacterial and fungal sources (see AAY06331-70).
 XX The sequence shows homology to T. reesei EgIII (see AAY06330). Also
 XX provided by the invention are vectors, host cells and methods
 XX for the recombinant production of such enzymes, which can be used
 XX in the treatment of cellulose-containing textiles, as feed
 XX additives, in the treatment of wood pulp, in the reduction of
 XX biomass to glucose, in the stone washing of indigo dyed denim, or
 XX as laundry detergent components (all claimed).
 XX Sequence 371 AA:
 SQ
 Query Match 11.5%; Score 635.5; DB 20; Length 371;
 Best Local Similarity 40.3%; Pred. No. 1.2e-32;
 Matches 149; Conservative 52; Mismatches 130; Indels 39; Gaps 15;
 QY 695 VAAVMAAGNTSAPSTPYDCTPPGPNQ-----GVTSDGEYRQTNEMNSSAQCCLTINT 749
 DB 16 lslalqalltaapag-----anqgldrygtltlqg-ryvvqnmrwgtsatqg--lnv 65
 QY 750 AMCAWVTSTANFSGGTGAPATYPSIYKGCWMGNCCTTKNVGMPIQISQISAVTSMTQ 809
 DB 66 tngvfeitqadqsvplngapksydvqchgnacprt-1pmrlslsapsvsvyry 124
 QY 810 VSSGAVDAVYDWTNSTPTTGTGPNGTETIMTWLNSRGVOPFGSQRTGTVAGHTWNNW 869
 DB 125 tngvvaaydvlwldptprtg-vnrteimwfnrvpvpqipsvgt-ahvgsrsewv 182
 QY 870 QGQTSWKIISYVLTFGATISINLDKATFADAARGSLNTSD-YLDDVEAGFEIWOGGQ 928
 DB 183 tngvngndvlsif-lapsalswsvfdvkd-fvdqavshglapdwyltalsigafepwegegt 240
 QY 929 GLGSNSFSVSVTSGG-----VACRAITYVNSDMGSGFATYVNTGSRATNGTWV 979

```

Db      241          |||||:::|||              ||::||::||||||:|||||::|||
           glavnsfasavnaggnngtgpqtpacqvsvstht-wpggffvdtltlnlgtptdpywel 299

QY      980 AMSREGNQTNTNNWTALTQSASSTYATNLVSN-NVIPGCGSTTGFNGSYSGTN-AAPT 1037
           ::|||::||::|||::|||::|||::|||::|||::|||::|||
Db      300 dftlpaghtvcaswallspasgavtlarslgsglaangtgtsfgfqtssgltfnapa 359

QY      1038 -----LST 1041
           |||
Db      360 ggrlngtscet 369

RESULT   8
AA014880
ID    AA014880 standard; Protein. 371 AA.
XX
AC    AAB014880;
XX
DT      21-NOV-2000 (first entry)
DE    Streptomyces lividans CelB EGIII-like cellulase.
XX
XX    Streptomyces lividans; Trichoderma reesei; endoglucanase III; EGIII;
KW    cellulase; mutant; enzyme stability; textile treatment;
RV    wood pulp treatment; feed additive; detergent.
XX
OS    Streptomyces lividans.
XX
WO200037614-A2.
PN      29-JUN-2000.
PD
XX
PF      12-NOV-1999; 99WO-US26704.
PR
XX      18-DEC-1998; 98US-0216295.
PA      (GENEV ) GENENCOR INT INC.
PI      Mitchinson C, Wendt DJ;
DR      WPI: 2000-482483/42.
PS
PT      Novel endoglucanase III or endoglucanase III-like cellulase useful for
PX      treating textiles and wood pulp comprises a substitution or deletion at
PY      specified positions in the wild form of endoglucanase III -
PZ
Example 1; Fig 3; 52pp; English.
XX
XX      The present sequence is a cellulase related to endoglucanase III (EGIII)
CC      from Trichoderma reesei. EgIII-like genes were isolated from genomic DNA
CC      libraries constructed from various microorganisms by PCR. The isolated
CC      libraries showed significant homology to EGIII from T. reesei. Certain
CC      substitution and deletion mutations have been incorporated into EGIII and
CC      EGIII-like cellulases to produce variant enzymes with improved stability,
CC      e.g., increased resistance to temperature stress. The mutants may be used
CC      in textile and wood pulp treatment, as a feed additive, and for reducing
CC      biomass to glucose. They are also useful for stonewashing or indigo dyed
CC      denim and as an agent in laundry and dish detergents.
XX
SO      Sequence 371 AA:

```

```

QY      VSSAGVAADIDIMNNSPTTGGDPNGNEIMILNNSRGVOPFGSCQATATVTAAGHTMNV  869
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      125 tgnvyvnaagidwdrprcng -vntelmwfnrvgpqrjpspygc -ahvgrstwew  182

QY      QGGQTSMKIISIVLTGPGATISNLDLKAIFADAAANGSLNTSD -YLLDVEAGFEIWMGQ  928
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      183 tgsnrgndvslf -lapaalaswafdvkd -fvdqavahglatpdywyltsiqagfepwegt  240

QY      GLGNSRFSVSATSGC-----VACRAIYYVNSDMWGSGFATVTVNTGSRATNMGWTV  979
      || |||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      241 glavnsfssavnaagvgngqf -lpgtpraaqqsylch -wpgyfdlttlrtgslpydvgwel  299

QY      AWFSGNQTVTNWNTALITQSGASVYATMLNLS -NVIQPGQSTTEGFENGSSYGIN -AAP  1037
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      300 dftlpaghtvtsamnalispasgavatarstgsngriaangtqsfqfsgtsgtgnapa  359

QY      1038 -----LSC  1041
      |||
Db      360 ggrlmgtsct  369

```

RESULT 9
 ID AAY84345
 AC AAY84345 standard; Protein; 371 AA.
 XX AAY84345;
 DT 12-JUL-2000 (first entry)
 XX
 DE Amino acid sequence of an endoglucanase III (EGIII)-like cellulase.
 XX
 KW Endoglucanase III; EGIII; EGIII-like cellulase; surfactant stability;
 KW cellulase; textile processing; textile cleaning; stonewashing;
 KW indigo dyed denim; cellulose containing fabric; fabric smoothness;
 KW pill removal; fibril removal; cotton; cellulosic fibre; dyeing; detergent;
 KW animal feed; wood pulp; paper; grain; biomass reduction; glucose.
 XX
 OS Actinomyces sp.
 XX
 PN WO200014208-A1.
 PD 16-MAR-2000.
 XX
 XX 24-AUG-1999; 99WO-US19154.
 XX
 XX 03-SEP-1998; 98US-0146729.
 PR
 XX
 PA (GENEV) GENENCOR INT INC.
 XX
 P1 Fowler T;
 P1
 DR WPI; 2000-271052/23.
 XX
 PT Novel variant endoglucanase IIT-like cellulases with improved
 PT surfactant stability and resistance to temperature stress, useful for
 PT textile processing or cleaning, treating wood pulp, food and grain, and
 PT reducing biomass to glucose -
 XX
 PS Disclosure; Page 64-65; 73pp; English.
 XX
 CC The present sequence represents an endoglucanase III (EGIII)-like
 CC cellulase. The cellulase has homology to the Trichoderma reesei EGIII
 CC protein. The variant cellulases have improved temperature stability,
 CC and improved surfactant stability. The variant cellulases and
 CC compositions containing them are used in textile processing or cleaning,
 CC e.g. stonewashing of indigo dyed denim, and modifying the texture, feel
 CC or appearance of cellulose containing fabrics (e.g. improving fabric
 CC smoothness or removing pills and fibrils). The compositions may also be
 CC used for the removal of immature or dead cotton from cellulosic fibres
 CC or fabric, which can cause uneven dyeing. The cellulase may also be used
 CC in a detergent composition for washing laundry and dishes and in the
 CC treatment of animal feed, wood pulp, paper, non-animal foods and grains.

the enzyme may also be used in the reduction of biomass to glucose

Sequence	371 AA
SQ	

Query Match	11.5%;	Score 635.5;	DB 21;	Length 371;
Best Local Similarity	40.3%;	Pred. No. 1.2e-32;		
Matches 149;	Conservative	52;	Mismatches 130;	Indels 39;
				Gaps 15

```

Oy      695 VAAVDAGNTSASTRPVDCPCPNON-----GVTSVODGEFRVOTNMENSSAOOCLfnt 749
Db      16 laaslgalltaaeapq-----anqlcdrytlltld-ryyvqnmrwgaatgc---inv 65
Oy      750 ATGAMTVSTANFSGGTGGARATPSLYKCGCHNGNCTKNWCMPIODISOIGSAVTSMTQ 809
Db      66 tgnqfeltqaagsvptrqapaksyysydgchyncaprrt-lpmrlssjapsavsry 124
Oy      810 VSSGAYDVAIYDWTNSTPTTGGPNGTEIWMILNSRGVQVPQSGSATGAVVYAGHTNMV 869
Db      125 tgnqyynaaydlwldprtrng-vnrteimlwfnrvgpvqjlspevgt-ahnygsrsew 182
Oy      870 OGQOQTSWKIISYLVLPKATISNNDKATPADAAAGSLNTSD-YLADVACFEIWOQGO 928
Db      183 tsgnsndatlef-lapsaaiswafdvrd-fvdqavahglatpdrwyltslqagfepwegt 240
Oy      929 GJGNSFESVYSGG-----VACRATVYNSDMGSGCFATVTVYNTGSRATNGMTV 979
Db      241 glavmfsfaavagngngtptgtapaaqvsyhtl-ypgqftvdcltlitlgsbpvwdg 299
Oy      980 AMSFGNGQTVYNTWNTALTQSGASTATNLXSN-NVYOPGOSTYFGNGSGSTN-AAPT 1037
Db      300 dfclpghvhetasamallspaaagavlatrsgnrlaangtqsfqglsagtgfnaba 359
Oy      1038 -----LSCF 1041
Db      360 ggrlngtsc 369

```

RESULT 10
AA84346
ID AA84346 standard; Protein: 429 AA.
XX
AC AA84346;
XX
DT 12-JUL-2000 (first entry)
DE
XX Amino acid sequence of an endoglucanase III (EGIII)-like cellulase.
XX
XX Endoglucanase III; EGIII; EGIII-like cellulase; surfactant stability;
KW cellulase; textile processing; textile cleaning; stonewashing;
KW indigo dyed denim; cellulose containing fabric; fabric smoothness;
KW pill removal; fibril removal; cotton; cellulosic fibre; dying; detergent
XX animal feed; wood pulp; paper; grain; biomass reduction; glucose.
XX
OS Streptomyces lividans.
XX
XX WO200014208-A1.
XX
XX 16-MAR-2000.
XX
XX 24-AUG-1999; 99WO-US19154.
XX
XX 03-SEP-1998; 98US-0146729.
XX
XX (GENB) GENENCO INT INC.
XX
XX
XX Fowler T;
XX
XX WPI: 2000-271052/23.
XX
XX Novel variant endoglucanase III-like cellulases with improved
XX surfactant stability and resistance to temperature stress, useful for
XX textile processing or cleaning, treating wood pulp, food and grain, and
XX

reducing biomass to glucose -
Disclosure; Page 65-66; 73pp; English

XX The present sequence represents an endoglucanase III (EGIII)-like
CC cellulase. The cellulase has homology to the *Trichoderma reesei* EGIII
CC cellulase. The variant cellulases have improved temperature stability,
CC protein. The variant cellulases have improved cellulases and
CC and improved surfactant stability. The variant cellulases and
CC compositions containing them are used in textile processing or cleaning,
CC e.g. stonewashing of indigo dyed denim, and modifying the texture, feel
CC or appearance of cellulose containing fabrics (e.g. improving fabric
CC smoothness or removing pills and fibrils). The compositions may also be
CC used for the removal of immature or dead cotton from cellulosic fibres
CC or fabric, which can cause uneven dyeing. The cellulase may also be used
CC in a detergent composition for washing laundry and dishes and in the
CC treatment of animal feed, wood pulp, paper, non-animal foods and grains
CC. The enzymes may also be used in the reduction of biomass to glucose.

Seq	Sequence	429 AA
1	...	
2	...	
3	...	
4	...	
5	...	
6	...	
7	...	
8	...	
9	...	
10	...	
11	...	
12	...	
13	...	
14	...	
15	...	
16	...	
17	...	
18	...	
19	...	
20	...	
21	...	
22	...	
23	...	
24	...	
25	...	
26	...	
27	...	
28	...	
29	...	
30	...	
31	...	
32	...	
33	...	
34	...	
35	...	
36	...	
37	...	
38	...	
39	...	
40	...	
41	...	
42	...	
43	...	
44	...	
45	...	
46	...	
47	...	
48	...	
49	...	
50	...	
51	...	
52	...	
53	...	
54	...	
55	...	
56	...	
57	...	
58	...	
59	...	
60	...	
61	...	
62	...	
63	...	
64	...	
65	...	
66	...	
67	...	
68	...	
69	...	
70	...	
71	...	
72	...	
73	...	
74	...	
75	...	
76	...	
77	...	
78	...	
79	...	
80	...	
81	...	
82	...	
83	...	
84	...	
85	...	
86	...	
87	...	
88	...	
89	...	
90	...	
91	...	
92	...	
93	...	
94	...	
95	...	
96	...	
97	...	
98	...	
99	...	
100	...	

Query Match	11.5%	Score 635	DB 21	Length 429
Best Local Similarity	39.5%	Pred. No. 1.5e-32		
Matches 153	Conservative 45	Mismatches 144	Indels 44	Gaps 12

[illegible]

RESULT	11
AAV08473	
ID	AAV08473 standard; Protein; 371 AA.
XX	
XX	
AC	AAV08473;
XX	
DT	28-JUL-1999 (first entry)
XX	
DE	Actinomycete sp. 36KD cellulase protein.
XX	
KM	Cellulase; detergent; animal feed; nutritional value; textile;
KM	stone washing; texture modification; appearance; cellulosic fabric;
KM	pulp; draining; paper; baking additive; starch treatment; grain;
KM	high-fructose corn syrup production; ethanol production; fibre reduction;
KM	milling.
OS	
XX	Actinomyces sp.
XX	
PN	WO9925846-A2.

XX 27-MAY-1999.
 PD 18-NOV-1998; 98WO-US24649.
 XX 24-JUN-1998; 98US-0104308.
 PR 19-NOV-1997; 97US-0974041.
 PR 19-NOV-1997; 97US-0974042.
 XX (GENV) GENENCOR INT INC.
 PA Jones BE, Van Der Kleij WAH, Van Solingen P, Weyler W;
 XX WPI: 1999-347481/29.
 DR N-PSDB; AA772359.
 XX
 PT New Actinomycete cellulase useful in detergent compositions, in
 PT animal feeds and in treatment of textiles
 XX
 PS Claim 1; Fig 1; 36pp; English.
 XX
 CC This invention describes a novel cellulase isolated from an Actinomycete
 CC sp. which can be used in detergent compositions, as animal feeds (to
 CC increase nutritional value) and in treatment of textiles (e.g. stone
 CC washing or modifying texture, feel and/or appearance of cellulosic
 CC fabrics, including removal of 'immature' or 'dead' cotton), pulp (to
 CC improve draining) and paper. They may also be used as baking additives,
 CC for treating starch (in production of high-fructose corn syrup or
 CC ethanol) and for treating grain (to reduce fibre during milling).
 XX
 SQ Sequence 371 AA:

Query Match 11.4%; Score 633.5; DB 20; Length 371;
 Best Local Similarity 40.0%; Pred. No. 1.6e-32;
 Matches 148; Conservative 52; Mismatches 131; Indels 39; Gaps 15;

QY 695 VAAVDAAGNTSAPSTPVDCTPGPNQ-----GVTSDQDEHYRQTNEMSSAQOCLTINT 749
 DB 16 laslglalllaaapq-----anqicdrlygtltqdr-ryvvnqrwtgsatgc--Inv 65
 QY 750 ANGAMTVSTANSGGTGAPATYPSIYKGCCHNGNCTTKVGNPQIQSGASAVTSMTQ 809
 DB 66 tgnngfeitqadgsvpngapksyvsydgchyncaprtt-lpmrlsissgsapsvsyry 124
 QY 810 VSSGAYDAVADYDWTNSTPTTGGPNCGETIMWINSRGVQPGSQATGVTVAAGHTWNV 869
 DB 125 tgnngvnaaydlwldprrtng-varteimlwfnrvqvpqjpspygt-ahvggrswev 182
 QY 870 QCGOQTSWKIISYVLPFGATISNLDKAFADAARGSLNTSD-YLLDVEAGFEIWOQGO 928
 DB 183 tsgngsndvlsf-lapsalswsfdvkd-fvdqavshglatpdywyltslqgfepeggt 240
 QY 929 GLGSNSFSVSYSGG-----VACRATYYVNSDMSGSGFATVTVTGTSGRATGWTY 979
 DB 241 glavnsfssavnaagngngtptgpaacqyshtc-wpgsfvtdtltntgstprvgwel 299
 QY 980 ANSFGNQTNTYNTWNTALQSGASVTATNLXSN-NVIOPGQSTTFGFGNSYSGTN-AAPT 1037
 DB 300 dflipaghtvsvwnallspasgavlarstgngriaangtqsfqgltsaggaftapa 359
 QY 1038 -----LSCGT 1041
 DB 360 garlntgstct 369

RESULT 12
 AA67496
 ID AA67496 standard; protein; 371 AA.
 AC AA67496;
 XX
 DT 19-MAY-2000 (first entry)

XX Actinomycetes cellulase protein sequence.
 DE Cellulase; Actinomycetes; detergent; feed additive; textile treatment;
 XX pulp; paper.
 KW Streptomyces sp.
 XX
 OS
 XX
 FH Key
 FT Peptide
 FT Protein
 FT Location/Qualifiers
 FT 1..27
 FT /note= "signal sequence"
 FT 28..372
 FT /note= "mature protein"
 PN WO200009707-A1.
 PD 24-FEB-2000.
 XX
 PF 28-MAY-1999; 99WO-US11971.
 XX
 PR 24-JUN-1998; 98US-0104308.
 PR 18-NOV-1998; 98WO-US24649.
 PR 28-MAY-1999; 99US-0321981.
 XX
 PA (GENV) GENENCOR INT INC.
 XX
 PI Jones BE, Van Der Kleij WAH, Van Solingen P, Weyler W;
 XX WPI: 2000-224344/19.
 DR N-PSDB; AA257029.
 XX

PT A novel Actinomycetes cellulase and related DNA, useful for detergent
 PT compositions, treating textiles and paper or pulp
 XX
 PS Claim 1; Fig 1; 72pp; English.
 XX

CC The invention provides a cellulase from Actinomycetes. The cellulase can
 CC be used in a detergent composition, as an additive for animal feed and
 CC for the treatment of textiles or pulp and paper. The DNA encoding the
 CC cellulase can be used to identify homologous cellulases and for
 CC recombinant production of cellulases. The present sequence represents a
 CC cellulase from Actinomycetes.
 XX
 SQ Sequence 371 AA:

Query Match 11.4%; Score 633.5; DB 21; Length 371;
 Best Local Similarity 40.0%; Pred. No. 1.6e-32;
 Matches 148; Conservative 52; Mismatches 131; Indels 39; Gaps 15;

QY 695 VAAVDAAGNTSAPSTPVDCTPGPNQ-----GVTSDQDEHYRQTNEMSSAQOCLTINT 749
 DB 16 laslglalllaaapq-----anqicdrlygtltqdr-ryvvnqrwtgsatgc--Inv 65
 QY 750 ANGAMTVSTANSGGTGAPATYPSIYKGCCHNGNCTTKVGNPQIQSGASAVTSMTQ 809
 DB 66 tgnngfeitqadgsvpngapksyvsydgchyncaprtt-lpmrlsissgsapsvsyry 124
 QY 810 VSSGAYDAVADYDWTNSTPTTGGPNCGETIMWINSRGVQPGSQATGVTVAAGHTWNV 869
 DB 125 tgnngvnaaydlwldprrtng-varteimlwfnrvqvpqjpspygt-ahvggrswev 182
 QY 870 QCGOQTSWKIISYVLPFGATISNLDKAFADAARGSLNTSD-YLLDVEAGFEIWOQGO 928
 DB 183 tsgngsndvlsf-lapsalswsfdvkd-fvdqavshglatpdywyltslqgfepeggt 240
 QY 929 GLGSNSFSVSYSGG-----VACRATYYVNSDMSGSGFATVTVTGTSGRATGWTY 979
 DB 241 glavnsfssavnaagngngtptgpaacqyshtc-wpgsfvtdtltntgstprvgwel 299
 QY 980 ANSFGNQTNTYNTWNTALQSGASVTATNLXSN-NVIOPGQSTTFGFGNSYSGTN-AAPT 1037
 DB 300 dflipaghtvsvwnallspasgavlarstgngriaangtqsfqgltsaggaftapa 359

QY	1038	-----L	SCT	1041
DB	360	garl	ngtact	369
QY	13	RESULT		
AA	84797	AA	84797	standard; Protein: 395 AA.
AC	AA	84797		
DF	08-AUG-2000		(first entry)	
DE	Amino acid sequence of a C1 endoglucanase polypeptide.			
KW	C1 endoglucanase; promoter; mutant; expression-regulating region; secretion signal sequence; enzyme; cellulase; xylanase; pectinase; lipase; protease.			
OS	Chrysosporium lucknowense.			
FH	Key	Location/Qualifiers		
FT	Peptide	1..28		
FT		/note="signal peptide"		
PN	WO200020555-A2.			
PD	13-APR-2000.			
XX	06-OCT-1999;	99WO-NL00618.		
XX	06-OCT-1998;	98WO-EP06496.		
XX	(AARL-) AARL INC.			
XX	Emailfarb MA, Burlingame RP, Olson PT, Sinitayn AP, Parriche M; Bousson JC, Pynnönen CM, Punt PJ, Van Zelfl CMJ;			
PI	WPI: 2000-317705/27.			
XX	N-PSDB: AAA14818.			
DR	Mutant Chrysosporium strain comprising nucleic acid encoding proteins, particularly industrially important enzymes such as cellulases, xylanases, pectinases, lipases and proteases			
XX	Dislosure; Page 43-46; 92pp; English.			
XX	The present sequence represents a Chrysosporium C1 endoglucanase. The polynucleotide sequence comprises a promoter sequence, and can be linked to a nucleic acid sequence of the invention to control expression. The specification describes a mutant Chrysosporium strain which comprises a nucleic acid sequence encoding a polypeptide of interest, where the nucleic acid sequence is linked to an expression-regulating region and a secretion signal sequence. The mutant strain is useful for producing heterologous proteins and polypeptides of interest, e.g. industrially useful enzymes such as cellulases, xylanases, pectinases, lipases and proteases.			
CC	Sequence	395 AA;		
QY	Query Match	10.4%; Score 574.5;	DB 21;	Length 395;
DB	Best Local Similarity	34.7%; Pred. No. 1e-28;		
Matches	160; Conservative	58; Mismatches	156; Indels	87; Gaps
QY	15 RCTTRMPAISKRLRAGVLAGAVSIAAIVPLAMQHPAIAATVHDNPYAGATFEVNPYMAQ	74		
DB	21 rttpqkpr--rqasg-casavtladas-----tnvfqgy---tlhpnmfya	60		
QY	75 EVOSEANQNTATLAAKRRVSTYSTAWMDXIAAINGVNGSGPGLTYYIDALSOQOQGT	134		
DB	61 evaaaeaaasdsdaaleakavadvgtflwldtlenlg-----rllepai--edvpc	108		

OY	135	PEVIEIVYIDLPGRDCAALASNGELPATLAAGLOTYETQYIDIPASILSNPKYSSLRIVTI	194
OY	136	PEVIEIVYIDLPGRDCAALASNGELPATLAAGLOTYETQYIDIPASILSNPKYSSLRIVTI	194
Db	109	enivglivlydIpgfrdoeakasngeI--kvgeIdrytleyIdkIaeI1k--abentafay	164
OY	195	IEPDSLPNNVMTMSIOACGTAFAVPYEOGIEVALTKLHAIPNNVIYMDAAHGILGMPNNA	254
Db	165	IepdaiplnIvtrnsdlqtccqsaasyreyayaIkqIin-IpnnvmyIdaIhgwlgywaol	223
OY	255	SGYVEOVOKVLNASISGVNIDGFEVTNTAVYTPLEKEPFMTATQOVGQPYESANFYOMNP	314
Db	224	kpgaegeIaevyIsaagspsqvgrIslnvagv-----nawdegepe-fsdaedagynkc	274
OY	315	IDEADYADVLYSRILVAAGFPSSIGMLIDPLRNGWGCPNEPTGSTATDVNTFVNQSKIDL	374
Db	275	geqkIyInIrfgeelksaemphn--aIIdtgrngvtI-----1	309
OY	375	RHNRGLMGNQNGAGCGPQOASPTDPPNNHLADYVWIKPPGESDGTSAASDPTTGKKSDP	434
Db	310	Idewgdcwcnvngagfgvtrptanqgd--eladaIvwwkpggesdgIsds-----aaryds	362
OY	435	MCDPTYTTSYGVLTNALPNSPIAGWEPFAQFDOLVANARPA	475
Db	363	fcgk-----pdafkpspsaegvtnqngyfemllInnaps	394
RESULT 14			
ID	AAV01077		
XX	AAV01077	standard; Protein; 473 AA.	
AC	AAV01077;		
DT	08-JUN-1999	(first entry)	
DE		H. insolens Cel6A fungal cellulase protien sequence.	
KW	Cel6B: Cel6A: fungal cellulase; cleaning composition; conditioner;		
KM	cellulolytically active protein; endo-1,4-beta-glucanase; enzyme;		
KW	fabric softener; Humicola-like cellulase; glycosyl hydrolase family 6;		
KM	detergent composition.		
OS	Humicola insolens.		
PN	MO9901544-A1.		
PD	14-JAN-1999.		
PF	02-JUL-1998;	98WO-DK00299.	
PR	04-JUL-1997;	97DK-0000813.	
PA	(NOVO) NOVO-NORDISK AS.		
PI	Andersen KV, Damgaard B, Lund H, Nielsen JB, Schuelein M;		
DR	WPI: 1999-106046/09.		
DR	N-PSDB: AAX27948, AAX27958.		
PT	Cleaning composition containing Humicola endo-beta-1,4-glucanase		
PT	useful as detergent compositions or additives, or as fabric		
PT	conditioners		
PS	Example 3; Page 243-245; 271pp; English.		
CC	This sequence is the Humicola insolens Cel6A fungal cellulase.		
CC	The invention relates to a cleaning composition (A) that contains at		
CC	least one enzyme (I) with cellulolytic activity and at least 25 wt. % of		
CC	total cellulolytically active protein present is Humicola		
CC	endo-1,4-beta-glucanase or a Humicola-like cellulase of the glycosyl		
CC	hydrolase family 6. (A) are used as detergent compositions or additives,		
CC	or as fabric softener or conditioner. (I) provides colour clarification,		
CC	and possibly soil removal, without significant weakening of cellulose		
CC	materials during pre-soaking or wet storage (contrast use of enzymes from		
CC			

Db 407 mriptant---ghqyvdafrwvkpggcdgts-----dttarydyhcg-----leda 450
Oy 451 LPNSPIAGOWFPQFDQLVANARP 474
| : | | | | | : | | |
Db 451 lkpeapeagwrfqayfegqlrnanp 474

Search completed: August 29, 2002, 16:21:46
Job time: 321 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:24:05 ; Search time 72.78 seconds
(without alignments)
1377.043 Million cell updates/sec

Title: US-09-917-384-6
 Perfect score: 5536
 Sequence: 1 MERTQSGNCRVQGTTRM.....GFGSYSGTNAAPTLSCTAS 1043

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

```

Post-processing: Minimum Match 0%
Maximum Match 100%

```
Database :
1: plr1: *
2: plr2: *
3: plr3: *
4: plr4: *
```

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1707	30.8	872	2	S49541	cellulase - Cellul
2	1239.5	22.4	683	2	A83704	1,4-beta-cellulobios
3	1051	19.0	576	2	T35270	probable secreted
4	1000	18.1	539	2	A55876	cellulose 1,4-beta
5	641	11.6	438	1	S70602	cellulose 1,4-beta
6	629.5	11.4	382	2	JCZ871	cellulase (EC 3.2.
7	586	10.6	1090	2	S59077	cellulose 1,4-beta
8	561.5	10.1	1045	2	A39199	cellulose 1,4-beta
9	548	9.9	471	1	A26160	cellulose 1,4-beta
10	548	9.9	471	1	A38979	cellulase (EC 3.2.
11	487	8.8	747	2	B47093	1,4-beta-glucanase
12	450.5	8.1	1711	2	T31303	cellulase (EC 3.2.
13	449.5	8.1	1742	2	T17120	mannan endo-1,4-be
14	434	7.8	1331	2	A48954	xymanase - Caldice
15	431.5	7.8	1779	2	T31085	alpha protein - C10
16	413	7.5	1854	2	S36859	cellulase (EC 3.2.
17	388.5	6.9	890	2	B41897	probable secreted
18	384.5	6.9	890	2	T35237	cellulase (EC 3.2.
19	383.5	6.9	1039	2	S02711	cellulase (EC 3.2.
20	383	6.9	499	2	A27198	cellulase (EC 3.2.
21	382	6.9	499	2	G69593	cellulase (EC 3.2.
22	381.5	6.9	499	2	JN0111	cellulase homolog
23	380	6.8	143	2	A41897	cellulase (EC 3.2.
24	379	6.8	508	2	A26874	cellulase (EC 3.2.
25	363.5	6.6	504	2	S34744	cellulase (EC 3.2.
26	355.5	6.4	380	2	T97847	cellulase (EC 3.2.
27	355.5	6.4	505	2	S39962	endo-glucanase - E
28	351.5	6.3	1162	2	T30433	scaffoldng prote
29	351	6.3	915	2	A43802	cellulase (EC 3.2

30	349	6.3	586	2	PG6006
31	339	6.1	1483	2	C97012
32	338.5	6.1	441	2	T12011
33	333	6.0	486	2	I40548
34	325.5	5.9	986	2	S12021
35	316	5.7	766	2	AB1452
36	315	5.7	639	2	BA3868
37	310	5.6	756	2	AB1088
38	309	5.6	1848	2	A44140
39	299	5.4	426	2	A42360
40	289	5.2	879	2	A47704
41	285.5	5.2	570	2	S56132
42	277.5	5.0	5188	2	B85547
43	277.5	5.0	5291	2	E90896
44	276	5.0	1230	2	S47466
45	275	5.0	488	2	A41961

ALIGNMENTS

```

RESULT      1
S49541
Cellulase - Cellulomonas fimi
C:Species: Cellulomonas fimi
C:Date: 29-Nov-1995 #sequence_rev:10n 01-Mar-1996 #extl_change 22-Oct-1999
C:Accession: S49541; A47093
R:Meinke, A.; Gilkes, N.R.; Kvan, E.; Kilburn, D.G.; Warren, R.A.V.; Miller Jr., R.C
Mol. Microbiol. 12, 413-422, 1994
A:Title: Cellulohydrolyase A (Cbha) from the cellulolytic bacterium Cellulomonas fimi
A:Reference number: S49541; MUID:94344030
A:Accession: S49541
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-872 <ME>
A:Cross-references: EMBL:L25809; NID:9456028; PID:AA036898.1; PID:9456029
R:Meinke, A.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.
J. Bacteriol. 175, 1910-1918, 1993
A:Title: Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase D (Cendb
A:Reference number: A47093; MUID:93209933
A:Accession: A47093
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 812-872 <ME>
A:Experimental source: ATCC 484
A:Note: sequence extracted from NCBI backbone (NCBIN:128120, NCBI:P:128121)
C:Superfamily: bacterial cellulose-binding domain homology
F:768-870/domain: bacterial cellulose-binding domain homology <BCB>
F:770-869/Distulfide bonds: #status Predicted

```

Query Match	30.88;	Score 1707;	DB 2;	length 872;
Best Local Similarity	38.66;	Pctd. No. 1.4e-84;		
Matches 403;	Conservative 123;	Mismatches 311;	Indels 206;	Gaps 27
Qy	23	ISKRLRAGLGNVSIASIVPLA - MCHPPIAATHVDNPNYAGATFFPNVNYMAOEVSEA 80		
Db	10	VRRRRA - VATAATATATLVAAPLTTLATASAAAPAHNDNPAGVAVQVNTPTMAASVMAA 68		
Qy	81	ANQT - NATIAAKRYSTSTAYMMDRTAAINGVNGSGGLTYYDAALSOOGT - PPEVI 138		
Db	69	GRSDAPALAAKRYVAGPTVMMDDRLSATINGDNGGLEFHDNDNVAAOKAAGVLPV 128		
Qy	139	EIYIYDLPGRDCALASNEELPATAGCLOTQYTDIPFASILSNPKYSRLIYTIIEPD 198		
Db	129	NVIYIYDLPGRDCCALASNEELPRITDAGLARKRYSEYIDIPDLADLPPEYSIRIATIEPD 188		
Qy	199	SLPNNAVYTNMSIOCATAVPVEYEGIEYALTKLHAIPNYIYVMDAHSGLMCPNNA --S 255		
Db	189	SLPPLTINISEPACQOAAAPYRGGVAYALDKHALIPNYNYIIDIGHSGMLGMDSGPSA 248		
Qy	256	GYVGQYKVLNASTIGVNGIDGVEYTNANTYPTLKEPMT - ATQGVGQPYESANFEVQMPD 314		

Db 249 TLEFAVAK--STTAGFASIDGFSDVANTTPEEPPLSSSLTINTPIRSSKFEWMD 306
 Qy 315 IDEADYAVDLVSRVLAAGPSSIGMLIDTLRWGCGPNEPTGPSTATDVNFVNSKIDL 374
 Db 307 FDEIDYTHMHLRLVLAAGPSSIGMLVDSRWGCGPNEPTGPSTATDVNFVNSKIDL 366
 Qy 375 RQHRGLMGNQAGGQPPQASPTDFPNHLDAYWIKRPGSDGTSASDPTTKKSP 434
 Db 367 RYHRGAWCPPLGAGIGRPEAPPSYASHDLAEFWIKRPGSDGTSASDPTTKKSP 426
 Qy 435 MCDPFTYTS--YGVLTNMLPNSPIAGOWEPAOFDOLVANARAVVSGGLKYKNNDSAP 492
 Db 427 MCDPFTYTS--YGVLTNMLPNSPIAGOWEPAOFDOLVANARAVVSGGLKYKNNDSAP 492
 Qy 493 GDNQKRP-----GLQVNTGSSVDLSVTYVRYWFTRDGSSSTLVYNCMDAAGCNGIRAS 548
 Db 476 VEDLAPVPVPGTGTAGTTATATSVPLS-----WTASTDVAATGYGVDYRGTLVGTATAT 529
 Qy 549 FGSVNPAPPTADTYQLSTGTGLAAGSGTGEIQNRVKNKSDMSNFTETDYSYGTNTTQ 608
 Db 530 STYVGTLP-ATAY--STYVAKADAGNVSA-----SAAAAATQ 567
 Qy 609 DWTKYTVYVNGVLVWGTEDVPPSVPTGLVYTVGSSSVSLAMNASTDNV---GVAHYNV 665
 Db 568 SGT-----VDTTAPSVPAAGLGTATTTTTPSLMTASTDAGSGVAGYEV 614
 Qy 666 YRNGVLVGPYVTSFTDGLAGTATYTVAAVDAAGNTSAPSTPYDCTPGPNONGVTSV 725
 Db 615 LKGTTVVGTATSYVTVGLTGTATYTVSVRAKDAVAGNTSASAASVAT----- 663
 Qy 726 QDGEYRVGTENMNSAQCCLTINTAGMTVSTANPSGSGTGAPATYPIYKCHMGNT 785
 Db 664 -----TOTGTVYVDTTA-----PSVPGTLAAGTT 687
 Qy 786 TKVNGMPIQISGASVTSWSTTOVSSGAYDVA-YDIWNTSPPTTGOPNGTEIMIMLNS 844
 Db 688 TSSVPL-----TWASTDNAGSGVAGYEVNGTTRVAT----- 721
 Qy 845 RCGVOPFGSQATGYTVAGHTNNWQGOOTSMKIIISYVLTLPATISLNDLKAIFADAAA 904
 Db 722 -----VTSSTYVTC-----LAADTAY 738
 Qy 905 RGSLSMTDYLLEVEAGEFEIMOGGGLGSNSFSVTSGGVACRATPVVNSDMGSGTAY 964
 Db 739 SFTYVAKDVAGNVSA-----SAANSARTQAATSGS--CTVYKISAS--NMTGTTGTV 787
 Qy 965 TWTNTGSRATNGVTAAVSPGNGQVTVNWTALTOGASVATNLSNNVIOGOSTTTPG 1024
 Db 788 EYKNNGTALNGLTIGFSPADGQKVSQGSAMESGSGTAVTAKNAPWNGTLAAGSVSISG 847
 Qy 1025 FNGSYSGTAAPT-----LSGT 1041
 Db 848 FNGTHNGTNTAPTATLNGVACT 870
 RESULT 2
 A82704
 1.4-beta-cellobiosidase XFI267 [Imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: A82704
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717
 A>Note: for a complete list of authors see reference number A59328 below
 A:Accession: A82704
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-683 <SIN>
 A:Cross-references: GB:A8003960; GB:A8003849; NID:g9106242; PIDN:AAF84076.1; GSPDB:GN001
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arida, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carier
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facinani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 J:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
 J.D.; Jungueira, M.L.; Kemper, E.L.; Klajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
 A:Authors: Martins, E.M.F.; Matsukuma, A.T.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
 , F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
 M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XFI267

Query Match 22.4%; Score 1239.5; DB 2; Length 683;

Best Local Similarity 31.9%; Pred. No. 1.8e-59;

Matches 314; Conservative 124; Mismatches 220; Indels 327; Gaps 27;

Qy 53 AATHYDNPAGATPFVNPYMAOEVOSEANOTNATLAAKMRVYTSYTAVMMDRIATAING 112
 Db 18 AEAHDNPEVDATSYLNPYSKIDIDTIAKNDVSLKEMQITKYPYGVWIDITDAKVG 77
 Qy 113 VNGCPG---LTTYDLAALSQOOGTPEVETIYIYDLPGRCAALASNGELPATAAGLOTY 169
 Db 78 GPKNGRLMLTGHDLAALAOKKGNPITATFVIYDIPGDCHALASNGELPLTPGLORY 137
 Qy 170 ETOYIDPIASLNSPKYSLRIYTIIEPDSLPNATYNNNSIOACATA--VPYEGDIEVAL 227
 Db 138 KREYIDTASISNPKYKIDIRIVNIIEPDSLNIINNSTPSCAKAKNTGIEBSIKTAL 197
 Qy 228 TKLHAIPNYIYMDAHSGLMCPNNASGYVOEVKVL-NASIGYNGIDGFEYNTANTYTP 286
 Db 198 NKLSEIPNYNNYMDIGHSGMLMDNRIIPAVSLYKIVISTAGASVNGFINTANTTP 257
 Qy 287 LKEPRK-TATQVGGOPVESANFYQWNPDIADYADVLSRLVAAGPSSIGMLIDTLR 345
 Db 258 LIEPNLPNDLILIGOPTRSSKFEWNRFEDEMDSETLYNDFVAAGMPSIGFIIDYGR 317
 Qy 346 NMGGPNEPTGSTATDVTFYVNSKIDLRHGRGLMONGAGLGLOPQASPTDPMNL 405
 Db 318 NMGGPNEPTGS-ATGNDVNSIYNSGRIDRNRHGMCKDAGIGLPIAT---PGGHV 372
 Qy 406 DAYVWIKPGESEDGTSASADPTTGKSKDPMDPTTYSYGVLTNLPNSPIAGOWEPAOF 465
 Db 373 DAFQWIKPGEYSDGSSSLIPNDGKGFDKCDPTFTTPGVLTGALPLADPLSGDFHNAOF 432
 Qy 466 DOLVANARPAVYSGGLKVOYKKNDSAPGDNQIKPGLQVNTGSSVDLSVTYVRYWFTRD 525
 Db 433 VRLINNAYPD-----SRTPLPG-----SGSSS-----S 456
 Qy 526 GGSSTLVYKNCMAAAGCNGIRASFGSVNPATPTADTYQLSTGTGLAAGSGTGEIQNRV 585
 Db 457 SGSS-----SSSGS-----SSSGSSSSSSSS----- 479
 Qy 586 NKSMSNFTETNDYSYGTNTFPQDMTKVTVYVNGVLVWGTEDVTPPSYVGLVTVGVSGS 645
 Db 480 -----SSGSSS-----SSSGSSSSSSSS-----SSSSS 492
 Qy 646 SYSLAMNASTDNVGAHVAVYVNRNGVLVGPYVTSFTDGLAAGTATYTVAAVDAAGNTS 705
 Db 493 SGSSSSSGSSSSSG-----SSSGSSSSSSSS-----SSSGSSSSS 515
 Qy 706 APSTPYDCTPGPNONGVTSVQDGEYRVGTENMNSAQCCLTINTAGMTVSTANPSGTT 765
 Db 516 SSGS-----SSSGSS 537
 Qy 766 GGAPATYPIYKCHMGNTKRVGMPIQISGASVTSWSTTOVSSGAYDVAVDIWTNS 825
 Db 538 S-----SSSGSS 558

```

RESULT      3
T35240
probable secreted cellulase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T35240
R:Seeger, K.U.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z1572
A:Accession: T35240
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-579 <SEP>
A:Cross-references: EMBL:AL031515; PIDN:CA20645.1; GSPDB:GN00070; SCOEDB:SC5C7.33
A:Experimental source: strain A5(2)
A:Genetics:
A:Gene: SCOEDB:SC5C7.33

```

[illegible]

R55976
 A55976
 cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) - Thermomonospora fusca
 C:Species: Thermomonospora fusca
 C:Date: 05-Jan-1996 #sequence-translation 05-Jan-1996 #text-change 22-Oct-1998
 C:Accession: A55976
 R:Zhang, S.; Lao, G.; Wilson, D.B.
 Biochemistry 34, 3386-3395, 1995
 A:Title: Characterization of a Thermomonospora fusca exocellulase.
 A:Reference number: A55976; PMID:95186436
 A:Accession: A55976
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-596 <ZNA>
 A:Cross-references: GB:U18978; NID:g664822; PIDD:AA67211.1; PID:g664823
 C:Specificity: bacterial cellulose-binding domain homology
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:40-140/Domin: bacterial cellulose-binding domain homology <BCB>
 F:41-139/Disulfide bonds: #status predicted

	Query Match	18.1%	Score 1000:	DB 2:	Length 596;
	Best Local Similarity	47.7%	Pred. No. 1.3e-46;		
	Matches 210;	Conservative 55;	Mismatches 143;	Indels 32;	Gaps
OY	50 PAIAATHVNDNPAAGATFEFNPYMAOEVSSEANQTNATLAKKMRVSTYSTAYVMRRIA				109
Db	172 PTMGEKVDPDFGAKALIVNPMWSKAAPGGS-----AVANESTAVMLDRIGA				221
OY	110 ING----VNGPGELTYLDLAISQQOGTPEVEIEIYVIDLGPRDCALASNGELPATNA				165
Db	222 IEGNDSEFTSGSMGRHLEAV-RQGSGPRLTYQVIITINLPORDCALASNGELGPDE-				278
OY	166 LOTETEOYIDPIASIILSN-PKYSLRIVTIIIEPDSLIPNAVTMSSIQACATVAPY-----				218
Db	279 LDKRKSETIIDPIDIMDFADIELRLVAILEIDSJPLNLVTYVINGNGGTELCAYKKONGG				338
OY	219 YEQGIETVALTLAHPNVITYINDAASHGMGLPNNASGYOEVQYULANS-IGVNGIDGE				277
Db	339 YVNGGVAYLRKIGELIPNYNTIDAHHGMWIGMSNEGPSVDIYEAAANMSGSTVDIVHGF				398
OY	278 VMTANTATPELKREPTTAQOYGQRPVESANYONPPDIEDADYADVILSRUYAAGEPSSI				337
Db	399 ISNTAYSATVPEDYDVNGTVNGOLTIROSKRWYDMNQYVELDSFYODLRQALLIKGRSDI				458
OY	338 GMLIDPLRMGSGMPNETGPTSTATDVNTFNOSKITDLQRHRGLMCNONGSGLQOPQAP				397
Db	459 GMILDIYSRKGMGSPNRPTSSSTSLNTLVYESRDIRIHFGNNCMNOAGAGLGERPTVYN				518
OY	398 TDEPNHAIDLAYVIKRPGESDGTSAASDETTGKSMDPCDCPTYYTSXGVLTN--ALPNS				454
Db	519 A----PGVDAYVWVKPPGEDSGASEEIPDNBEOKGFDRCMDPTYOGNARNGNPNPGALPNA				574
OY	455 PIAGOWFPAPQDQLVANARP 474				
Db	575 PISGHMFSAQFERELLANAP 594				

```

RESULT      5
S70602
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) II precursor - cultivated mushroom
N:Alternate names: cellulase
C:Species: Agaricus bisporus (cultivated mushroom)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: S70602
R:Yagoue, E.; Chow, C.M.; Challen, M.P.; Thuston, C.F.
Curr. Genet. 30, 56-61, 1996
A:Title: Correlation of exons with functional domains and folding regions in a cellul
A:Reference number: S70602; MUID:96269930
A:Accession: S70602
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-438 <YAG>

```

[illegible]

	Query Match	11.4%	Score 629.5;	DB 2;	Length 382;
	Best Local Similarity	39.9%;	Pred. No.	7.le-27;	Mismatches 147; Conservative 48; Mismatches 148; Indels 25; Gaps 9;
OY	LAAGTATVTVTAADVDAACGNTSAPSTPYDCTCPGPQNGVTYSVDGEYRVOTENNSSAQOC	744			
Dd	LRAVSALLTLFALMALLLLTATPAQAQDTTI-CPEFGSVIQQ-GRYVVQNRRMGTSAQC	68			
OY	LTINTATAMTVSTRANESGGTGGAIPATPSLYKCGHMCNCTTKAVMPAIIQSIGSVAITS	804			
Dd	VTT--ATDSGFRTQTADGSVPNPINGAKRSPVFYNCHYTNC-SPTALPARISGISPASS	125			
OY	WSTIQVSSGADVADVADIWTNSTPPTTGOPONTEIMINLSHGVOQPGRSQOATGVTAQH	864			
Dd	IISGYDVNAVNYTNASDIHTLDPTRTDG-VANTEIMINFNRGOIQPGSOVCIT-ASYAGR	183			
OY	TWNWVGOGOOTSMKIISTYYLVLPBGATSINSLDKAIFADAANGSLNTPSYDLLDYVEAGEFIW	924			
Dd	TWEWTSYGSGGNIDNVLSFY-APSAMSMSMGFDVMDFVRATVANGLANDMWYLNSIOAGEEPW	242			
OY	QGCGCLGSNSTSVYSTSG-----VACRAYTYVNSDMGSCFTATVTVN	968			
Dd	QNGAGLANVSSFSSTVNTGSGSONPDPPNGPCGPCPATCYSTATVNW-WPGGFTANVTVN	301			
OY	TGSRATNGTIVAAISPFCNQGTNYTWMTALTOSGASVTFNTLIYSNNVLTOPGGSTFEFNCS	1028			
Dd	NGBAPVDWMRLAFILTPSQSVYHAMNASVSPSSCAVATATGAEASRKRIAGAASSOFEGOA	361			
OY	YSGTNAAP	1036			
Dd	YSGSFAPQ	369			
RESULT					
S59077	Cellobiose 1,4-beta-cellubiosidase (EC 3.2.1.91) precursor - Cellulomonas fimi				
N:	Alternate names: exo-Beta-1,4-cellobiohydrolase				
C:	Species: Cellulomonas fimi				
G:	Date: 15-Feb-1996 #sequence revision 23-Feb-1996 #text_change 22-Oct-1999				
A:	Accession: S59077; S59044; I40694				
R:	Shen, H.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.J.				
B:	Biochem. J. 311, 67-74, 1995				
A:	Title: Cellobiohydrolase B, a second exo-cellobiohydrolase from the cellulolytic ba				
A:	Reference number: S59044; MWID:96003898				
A:	Accession: S59077				
A:	Molecule type: DNA				
A:	Residues: 1-1090 <SHED>				
A:	Cross-references: EMBL:L398827; NID:g790696; PID:NAB00822.1; PID:g790698				
A:	Accession: S59044				
A:	Molecule type: protein				
A:	Residues: 54-78;456-461 <SHW>				
R:	Shen, H.; Tonne, P.; Melnie, A.; Gilkes, N.R.; Kilburn, D.G.; Warren, R.A.; Miller,				
B:	Biochem. Biophys. Res. Commun. 199, 1223-1228, 1994				
A:	Title: Stereochemical course of hydrolysis catalysed by Cellulomonas fimi CenE, a m				
A:	Reference number: I40694; MWID:94197708				
A:	Accession: I40694				
A:	Status: preliminary; translated from GB/EMBL/DDBJ				
A:	Molecule type: DNA				
A:	Residues: 54-106 <RES>				
A:	Cross-references: GI:I29042; NID:g556035; PIDN:AAA50257.1; PID:g556036				
C:	Genetics:				
A:	Gene: cbnE				
C:	Superfamily: bacterial cellobiose-binding domain homology				
C:	Keywords: glycosidase; hydrolase; polysaccharide degradation				
F:	1-53/Domains: signal sequence #status predicted <SIG>				
F:	354-1090/Product: cellobiose 1,4-beta-cellubiosidase #status experimental <MAT>				
F:	789-1090/Domin: bacterial cellobiose-binding domain homology <BCB>				
F:	/g90-1089/Disulfide bonds: #status predicted				
Query Match	10.6%; Score 586; DB 2; Length 1090;				
Best Local Similarity	24.0%; Pred. NO. 6./e-24;				

Matches 299: Conservative 147; Mismatches 419; Indels 382; Gaps 62;

20 MPALSKRLRAGLAVASIASIVLMAQHPALATHVDNPFAGATFEVVPYMAQVQSE 79
 1 MSSTRRRSAMAAATVGV-SSFLAVAGITPAIAAGAGOP-AT-----VYTP 46
 80 AANQATNATLAAR-MRVSTSTAVMORIRAIANGVNGPGLTYLDAALSOOGTTPPV 137
 47 AASPVAADVGEYAOEFLAOY-----DKIK--DPANG-----YFSA-----OGIPYHA 87
 138 IEVYIDLP--GRDCAALASN-----GELPATAAGL-QTYET-----OYIDPIA 178
 88 VETLMEAPYGHETTSSEAVSWMLLEALYGVQVODMAPLNMHTMEKMYIPQSVQDP 147
 179 SILSNPYSILRIYITIEPDSLPNAVYN-----MSIOCATAVPYEGIEYALTKL 230
 148 NSFYNPSPATYAFEEFNHPSYPSOLNSGISGTDPIGALMAT-----YGNADVQMHML 203
 231 HAIPRVYLYMDAASG-WLGMPPNNAAGYVOEOKVLNAST-----GVNGI 274
 204 ADVDNITGFAGTTPGAGCTIGTATGTFINTFQNGPQESWETVPQSCFEKRYGKNKY 263
 275 DEFVNTANTYPLKEPFTATQOYGGPQVES---ANFYQNNPDIDEA----- 318
 264 LDEFTKDAASYA---KOMKYSASADARAVEAVYMAN--QVATEQCKADVAATVAKAAAK 319
 319 -DY---AVDLYSRLV-----AAGFPSSIGMLIDILRNMGWGPNE----- 353
 320 GDLRLTLEDKYFKKIGCTSPCAAGOREAAHYLLMTYMAAGANTDTSSGMARIGSSH 379
 354 -----PTGPTATDVNTFVNQSKIDLROHRLGKMONAGL--- 389
 380 AHFGYONPLAALSTDPKLTSPKAKADMAASM-OROLEFY-----TWLOASNGIAG 434
 390 -----GQPPQASPTDFPNALHDAVYIKPPGES-DGTSA-----ASDPTT 428
 435 ATNSMDGAYVQAPGAPTFEYGMCTEAPYVDPDSNRHMGOMAGVORVAYELYSAGNAQ 494
 429 GKSDPKDPTTYTSYGLVNLALPNSFIAGOMPFAFDOLVANARAVYSGGLKYOYKN- 487
 495 AAK-----ILDKWVPVAVVAVANSTDGASWKPKPSELKNTGKPD 530
 488 --NDSAPGDNDQIKPGLQI-VNTGSSSYDLSGVTVR---YWFTRDGGST---LVYKCD 536
 531 TNNAAAPTCGN--PGLTYEVTSYGDDVGAADTFRALLFFAAKSGDASDKAKALDAI 587
 537 WAAMCGNTRBASFGSVNPATPTADTYLQLSFT-----GGTLAAGSTGEHON-RVKNKSDW 590
 588 WA-----NNODPLGV--SAVETRGRDYKRPDDTYVANGDGIYIPSGWTGTPMNGDVIKPGV 640
 591 SNFETENDSYGTNTTPOWTKVTVYVNG-----VLVWG----- 624
 641 SPFLDISFYKKDPN-----MSKVOTFLDGGALPEQFRHRENAQTVAAALADYARLFDG 695
 625 --TEDVTPSPVPTGLVTVGVSSSVLANNASTDNVGAHVAVVYVNGVLVGOPTVYSTFD 682
 696 TTTDDTAPVPTPGLAGVYVSTEATISWTASTDTRVYGIVYGAIRKVGATATSTTD 755
 683 TGLAAGTAYTYTVAADAAGTSAFS-----TPVDCTPGPNONGYTSVQDEYVAVQ 733
 756 TGLTFASTAYATVRAFADAGVNSAPSAALVTTTKATPSPD-TTAPSPATIS----- 805
 734 TNEWSSAOCCLITNTATGANTVSTANFSGGTGGAATATYSIKGGCHMCNCTTKNVMPI 793
 806 -----SSS-----TANSYVIGMSASTDN-AGSGG--LAGY-DVYRGA-----T 839
 794 QISQISANTSMSTQVSSGAY-----DVAYDI-----WTNSTPTTGGQ-P 833
 840 RVAQT-TALTFTDGLITASTAIEYVBARDAVAGVNSAPSTAVSTKSDTTPOTAPSPV 898
 834 NGTEIMWLNRSRGVQFPGSOTATGVTYAGHTMVMWQOQTSWKLISVILPGATSIISNL 893
 899 AGLAAM-----TVEETSYA-LTMN-----ASTDYGGS 924

894 DLKA--IFADPAARGSLNNTSDYL---LDVEAGEF---IMOGGOGIGSNSFSVSYS- 941
 925 GLKGYDYVRGATRVGSGSTTASDYDTGLTAATAYQYTVATRDAGAVNSAASALSTTKTP 984
 942 -GGYACATRYVNSDMGSGFTATVTVTNGSRATNGVYASFSGGNQVYTNVMTALQOS 1000
 985 QVGGSCSVAYNAASS-WNSGFTASVRITNGTITINMSLGFLDLAGQVQGSMTWQOS 1043
 1001 CASVATNLYSNMYIOPGOSTTEFGNSYSGTNAAPT-----LSCT 1041
 1044 GSTVATATNAPMNGTILAPQVADVGFNGSHGQNPNSASTLNASCT 1090

RESULT 8
 A39199
 endoglucanase B (EC 3.2.1.-) - Cellulomonas flm1
 C:Species: Cellulomonas flm1
 C:Date: 08-Nov-1991 #sequence_revision 08-Nov-1991 #text_change 22-Oct-1999
 C:Accession: A39199
 R:Meinke, A.; Braun, C.; Gilles, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.J.
 J. Bacteriol. 173, 308-314, 1991
 A:Title: Unusual sequence organization in Cemb, an inverting endoglucanase from Cellu
 A:Reference number: A39199; MUID:91100298
 A:Accession: A39199
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1045 <MEI>
 A:Cross-references: GB:64644; NID:q144415; PID:AAA23086.1; PID:q144416
 C:Superfamily: fibronectin type III repeat homology; bacterial cellulose-binding doma
 C:Keywords: glycosidase; hydrolase
 E:945-1045/domain: bacterial cellulose-binding domain homology <RCB>
 E:946-1044/disulfide bonds: #status predicted

Query Match 10.1%; Score 561.5; DB 2; Length 1045;
 Best local similarity 26.7%; Pred. No. 1.3e-22;
 Matches 265; Conservative 105; Mismatches 382; Indels 241; Gaps 47;

146 PGGRD-----CALASN-----GELPATAAGL-----QTYEQYIDPIASISNP 184
 181 PGSDVAETRAALASAIYLVKDDPAVYATLVSHAKOLYTRADTYRGAISDCTAASAVY 240
 185 K-YSSLRIYITIEPDSLPNAVNTMSIOACATVAPYEGIEYALTKLAIPVYVYMDA 243
 241 KSMGYODELWCAVYLYKATGATYLAKEA-EYDKLGTENQST-----TRS 287
 244 HSGWLGMPPNNAAG-----YQDEVOKVINA-SIGVNGIDGFTVNTANTPLKE 289
 288 YKWTIANDNKOFGYVALLAMETGOKRYVDANRWLDYTVGVNG-----QKVPYSP--- 338
 290 PPMATQOYGGOPV-----ESAN-----FYQNMFD-----IDEDYAV- 322
 339 -----GGQAVLDSNGALRYAANTSFVALYSDMNTDATRKARYHDFGROINVALG 389
 323 -DLYSRLVYAGFPSSIGMLIDTLANGCGPNEPSTATDVNTFVNQSKIDLROHRLGM 381
 390 DNPRESSYVVG-----GANPPIAHRHRAHSMWLSITTPQASHVLY 433
 382 CNONGAGLGOPQASPTDFPNALHDA---YVWIKPGESEDCTSAASPTGKXSDMCPD 438
 434 ---GALVGGP--GSPND--AYTDSRDYV-----ANEVAYDYNAG----- 466
 439 TYTTSYGLVNLALPNSFIAGOMPFAO-----FDOLVANARAVYSGGLKYOYKNNDAP 492
 467 -FTSALARLVEEYGGPPLASFPPEPDPODDOLFVBAMLNOPSGFTFVKMKINOSAFP 525
 493 GDNOIKPGLQLVWNGSSVDLSVTYVYVFTPDG--GSSTLYVNCWMAAGCGNTRASFSG 551
 526 ARS-----LKNNAKRYWFTTDGEFASVYTLISANYSE--CG---AQSCK 563
 552 VNPATPTADTYLQLSFTGTLAAGST---GEIONRV-NKSDMSNFTENDSYG--TNT 605

[illegible]

RESULT 10
A38979
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) II - fungus (*Trichoderma viride*)
C:Species: *Trichoderma viride*
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A38979
R:Wang, J.R.; Zhang, M.F.; Huang, T.
Acta Genet. Sin. 22, 74-80, 1995
A:Title: The primary structure of cellobiohydrazase gene (CBH II) from *Trichoderma viride*
A:Reference number: A38979
A:Accession: A38979
A:Molecule type: DNA
A:Residues: 1-471 <MAN>
C:Genetics:
A:Gene: cbhII
A:Introns: 31/2; 160/2; 243/1
C:Superfamily: cellulose 1,4-beta-cellobiosidase II; fungal cellulose-binding domain
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:31-62/Domain: fungal cellulose-binding domain homology <FCB>

[illegible]

RESULT 11
B47093
cellulase (EC 3.2.1.4) Cend - Cellulomonas fimi
N/Alternate names: endo-1,4-beta-glucanase
C/Species: Cellulomonas fimi
C/Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 06-Dec-1996
C/Accession: B47093
J:Meinke, A.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.
J. Bacteriol. 175, 1910-1918, 1993
A:Title: Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase D (Cend),
A:Reference number: A47093; PMID:9320993
A:Accession: B47093
A:Status: preliminary
A:Molecule type: DNA; protein
A:Residues: 1-747 <MEI>
A:Experimental source: ATCC 484
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBI:128120, NCBI:128122)
C/Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Superfamily: bacterial cellulose-binding domain homology
C:Keywords: glycosidase; hydrolyase; polysaccharide degradation
F:044-744/domain: bacterial cellulose-binding domain homology <BCB>
F:045-744/disulfide bonds: #status predicted

Query Match	8.88;	Score 487;	DB 2;	Length 747;
Best Local Similarity	27.18;	Pred. No. 8.9e-19;		
Matches 168;	Conservative	67;	Mismatches 175;	Indels 210; Gaps 23;
QY	498	KPGIOLVNTGSSS-----	-VDLST-----	-YTVRYEPTRD 525
		: :		: :
Db	260	KRPVPTSTGGLRDYYGTWGGNLRCGRDHPIDLGADQOLYSPHDYGLPYFDQKFMQKD		319
QY	526	GGSTL-----VYNDDMAAMCGNIR-----	ASFGSVNRPATPRTAYL-----	-QLSFT 568
		::: : :		: : :
Db	320	FDKRSLADYVWGNMFLFHNDEDIAPLLIGEMGRLGODPRDCKMAALRLDVAERILSQT		379
QY	569	GGTL-AAAGSTGELQNRVKKSDMSNTELENDYSGTNTTFDQMTKVYV-----	NG	619
		: :		:
Db	380	FWVILNPNSGTGGLL-----LDPMKTDYV-----	KYSMLPEPTLMKHGCRYVGLDHOVPLGG	432
QY	620	V-----LYWGTEDVTPSPVPPGLVYVTVGSQSVSLANASTD-----	NGVAVHANYR	667
		: : :		:
Db	433	VGSFTGTSISQVGGGTEDTAPAPGLRAGPTASTVPLTMSASTDGGSSVAGYEYR		492
QY	668	NCGLVAGQPTVVSFTDGLAGTAYTYVVAADAAGNTSAPSPVDCPTPQPNQNGTVSQD		727
		: : : :		: :
Db	493	GTLVGTATTATSYVTGLADSAATFESVRKDKAGNTSA-----		531

RESULT 12
T31337
1,4-beta-glucanase (EC 3.2.1.-) - Anaerocellum thermophilum (fragment)
C:Species: Anaerocellum thermophilum
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T31337
R:Zverlov, V., Mahr, S., Riedel, K., Bronnenmeier, K.
Microbiology 144, 457-465, 1998
A:title: Properties and gene structure of a bifunctional cellulolytic enzyme
omains.
A:Reference number: Z21003; MUID:98154434
A:Accession: T31337
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1711 <2VE>
A:Cross-references: EMBL.Z86105; NID:e1071329; PID:e350354; PIDN:CAB06786.1
C:Genetics:
A:gene: celsA
C:Keywords: glycosidase; hydrolase

[illegible]

```

QY 319 D-----YAVDLYSLVAAGF-PSSIGMLIDLNRGCGPNEPTCP----- 357
Db 483 SGWPAKATDKLRFYVDL-SELIRAGYSPNOL-----TLSTNYOGAAGSGSPYWDASK 536
QY 358 -----STATDV-----NRFVNOSKIDLRQHRGL-WCNONG-----AGCGPPQASP 397
Db 557 NIYYLLVDFTGLIRPGODKKKEVQRIAPQVONQMDNSDYFQDQKGVSSSVYKT 596
QY 398 TDFPNAHLDAVYVIRKPGESDGT-----SAASDPYTKKSDPMCDPTYT----- 441
Db 597 KYIPLYDDVKKWGDGPGSGATPPTATATPTPTPTPTPTPTPTPTPTPTPTPTPTPT 656
QY 442 -----TSYGVLTNALPNSPIAGQWPPAOPDOLVANAARAVYGGGLAKYQKND 489
Db 657 TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 696
QY 490 SAPGDNQIKPGIQLVNTGSSSYDLSTVYRYWFTRDGSSSTLVYNCDMAACGNIRASF 549
Db 697 TNSTNTIRPMUKVYNTGSSSIDLSRVTIRYTYVDGKAQSAIS-DMAQIGASNTYKFX 755
QY 550 GSVNATPTADTYIQLSFTGT--LAAGSGTEIONRVKNSDMSNFTETNDYSY-GTNTT 606
Db 756 VLSSSVSGADYLLIEIGFKSGAGOLQAGKDTGEIQRNKSDMSNYNQNDMSMQSMTN 815
QY 607 FQDMTKVYVYVGVVWGTEDVTPPSVPTGLVYTVGSSSVSLANASTDNVCAHYANY 666
Db 816 YENKRVATYIDGLVWGOE-----PSGAPPT----- 842
QY 667 RAGVLVGOPTVTSFDTGLAAGTATYTYVAADAGNTSAPSTPYDCPTGPNONGVTSYQ 726
Db 843 -----PTAT-----PAPYTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 863
QY 727 DGEYRQINEMSSAQQCLTINTATGAMTVSTANSGTGAPATYPSIYKCHMGNCYT 786
Db 864 -----TATPTATPT-----PTPTPT----- 877
QY 787 KAVGMPRIQISQISAVTSMSTQVSSGADVAY-DIWTNSTPTTQGPNGTEIMITLNR 845
Db 878 -----SSSTPVAGQIKVLYANKETNSTNT-----TRPMLK----- 908
QY 846 GGVQPFQSGTATGTVVAGHTVWVWMOGOOTSMKLIISVLTLPGATSIINLDAKAFADAAR 905
Db 909 --VVNTGSSSIDLSRVTIRYTYVDGKAQSAISDMAQI-GASNTY-----FKFKVKS 958
QY 906 GSLMTSDVTLIDVAGEFEIWMQGGJGJGNSFSVTSVGVACRATYVYVNSDMWGSFTATYT 965
Db 959 SSVSGADYLLIEI--GFKSGAGOLQAGKDTGEIQRNK-----KSDMSN----- 999
QY 966 VNTGSRATNGVTFVMSFGNCTVNTNYMTALTQSGASYATATNLISNNYIQGGSTTTCF 1025
Db 1000 -YNOG---NDW---SMM-----OSMTNY-----GENVKVT-AVTDGLVWGOEP--- 1035
QY 1026 NGSYSGTNAAPT 1037
Db 1036 -----SGATPTPT 1043

```

RESULT 13
 cellulase (EC 3.2.1.-) precursor, thermoactive - Caldocellum saccharolyticum
 C:Species: Caldocellum saccharolyticum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
 C:Accession: T17120; A43745
 R:Te'o, V.S.; Sauli, D.J.; Bergquist, P.L.
 Appl. Microbiol. Biotechnol. 43, 291-296, 1995
 A:Title: Cella, another gene coding for a multidomain cellulase from the extreme thermoph
 A:Reference number: Z18698; MUID:95336703
 A:Accession: T17120
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1742 <TRQ>
 A:Cross-references: EMBL:L32742; NID:g5537499; PID:g537500; PIDN:AAA91086.1

R:luethl, E.; Bhana Jsmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
 Appl. Environ. Microbiol. 57, 694-700, 1991
 A:Title: Cloning, sequence analysis, and expression in *Escherichia coli* of a gene cod
 A:Reference number: A43745; MUID:91247819
 A:Accession: A43745
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1516-1544, 'A', 1546-1742 <LUE>
 C:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72860.1; PID:g144293
 C:Genetics:
 A:Gene: cels
 C:Keywords: glycosidase; hydrolase

Query Match 8.1%; Score 449.5; DB 2; Length 1742;
 Best Local Similarity 21.5%; Pred. No. 3e-16;
 Matches 249; Conservative 114; Mismatches 389; Indels 407; Gaps 48;

```

QY 29 AGVLGAVSIASIVPLANOHAIAATHV-----DNPTAGATFEVNPY-- 71
Db 173 AVVAETASLAAASIVLKDRNPTKATYIQAHAADLYEFAEVTKSDSGYTAANGYNSWSG 232
QY 72 -----MAQVQSEANQTNATLAAKMRVSTYSTAWMMDRTAATNGVGGGLFTYIDA 125
Db 233 FYDELSSMA-----AVWL--YLATND-----STYLRK 256
QY 126 ALSQOQGTPEVIEIYIDLPRDCALASNGE--LPATAGIQETETQYIDPIASISLN 183
Db 257 AESYVQ--NMPKISGSNIIDYKWAHCMDVHNGAALLAKITQDKYK-QIESHLDYMT 314
QY 184 P-----KSSIRIYTIIEPDSLPANAVTMSTI-----QACATAV-----PYEQGLE 225
Db 315 GYGERIKYTPGKLMWLDQMGSLRATYTAFLAFYSDWSGCPGTEKETEYRREFGSOIDY 374
QY 226 AL-----TKHAIPNYIYMDAAHSGM-----LGMPNNSAGYQ 259
Db 375 ALGSGRSPVVGCGTNPFRPHHRTAHSMAQSOSIPSHRRTLGALVGGGSDSDSYTD 434
QY 260 EVQKYLNASIGVNGIDFVYNTANTY-----TPLEPEM----- 292
Db 435 DISNRYNNEVACDYNAGFALAKMYLTYGNPDPFKAIETPTNDEFVEAGINASGTN 494
QY 293 -----TATQOVGQGPVESAIFQWNPDIIDEADYANDVLSRLVAAGF-PSSIGMLIDLNR 346
Db 495 FEIKAIYNQSGWPARATNKLFR-----YFVDL-SELIRAGYSPNOL-----TLST 541
QY 347 GNGGPEPTGP-----STATDV-----NRFVNOSKIDLRQHRGL-WCNON 385
Db 542 NYNQGAKVSGPYVWDBSSRNIIYIINDFTGLIRPGODKKKEVQRIAPQVONQMDNSN 601
QY 386 G-----AGLGOPPOASPTDFPNAHLDAVYVIRKPGESDGTSAAS-DPTTGKKSDDPMD 438
Db 602 DYSFQDIKGVSSGSYVKTXYIPLYDEDIKW-----GEEPSTGSVSPTPASVTPPTPTPT 657
QY 439 TYTTSYGVLTNALPNSPIAGQWPPAOPDOLVANAAR-----VYSGGLAKYQKND 491
Db 658 TATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 717
QY 492 PGDNQIKPGIQLVNTGSSSYDLSTVYRYWFTRDGSSSTLVYNCDMAACGNIRASFS 551
Db 718 STTNTIRPMUKVYNTGSSSIDLSRVTIRYTYVDGKAQSAIS-DMAQIGASNTYKFX 776
QY 552 VNPATPTADTYIQLSFTGT--LAAGSGTEIONRVKNSDMSNFTETNDYSY-GTNTT 608
Db 777 LSSSVSGADYLLIEIGFKSGAGOLQAGKDTGEIQRNKSDMSNYNQNDMSMQSMTN 815
QY 609 DMTKTVYVYVGVVWGTEDVTPPSVPTGLVYTVGSSSVSLANASTDNVCAHYANYRN 668
Db 837 ENEKYATYIDGLVWGOE-----PSGTPPA----- 861
QY 669 GVLVGOPTVTSFDTGLAAGTATYTYVAADAGNTSAPSTPYDCPTGPNONGVTSYDQ 728
Db 862 -----PTSTP-----TYVTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 899

```


Best local similarity 22.7%; Pred. No. 2.9e-15;
Matches 211; Conservative 105; Mismatches 325; Indels 287; Gaps 40;

```
QY 44 PLAMQHPAIAATHVDNPNY--AGATFEV--NPYMAOEVOSEANQTNAT--LAARKRVYS 96
Db 695 PTPPTPTVVSATPTPAPASAGSYWMPSESYGALKVYANGNMSSTTNVLPNKIKIEN 754
QY 97 TYSTAVMMDRIA-----AINGVNGGPGLTYYIDAA--LSQOQGTPEVIEIYIDLPGRD 149
Db 755 VGTAVVDLSRVKRYWYIDG-----EAAQSVSVASSINPAYIDVRV----- 797
QY 150 CAALASNGELPTAAGLQTY--ETQYIDPIASILSNPKYSLSRIYTLIEPDSLPMNAVTNMS 208
Db 798 -----KLGANAGADYYVEVGFSGAGVLAAGQSTKEIR-----LS 833
QY 209 IQACATAVPYEEOGIEYALTKLHAIPNYI-----YMDAHSGLGWPNN----- 253
Db 834 IOKSSGS---YQNSNDYSVRS---ANSYIENEKVYIDVY---LVMGREPGRNAQIK 882
QY 254 ---AGYVQEVQKVLNAGISGVNGIDGPTNTANTPTLKEPEMTAQOYGGQPYESANFYQ 310
Db 883 WYTAGNLSSPPTNVLNPKIKIENVG--TTAVDLSRVKRYWYIIDGEATQSVVAS--S 937
QY 311 WNP-----DIDEDAYAVDLSRLVAAGFPSSIGMLIDTLRNGMGPNNEPTGPS 358
Db 938 INPATIDRVKVLGAMAGADYYEV-----GKSGAGVL-----AAGQS 977
QY 359 TATDVNTFVNOSKIDLRQHRGLMCNONGAGLGQPPQASPTDFPNMHL--DAYVWIKPGES 417
Db 978 T-----KEIRLSIQKSSGSYNQNSNDYSVRSANSYIENEKVYIDVYLVWKEP--S 1027
QY 418 DGTSAASDPTTGKSDPMQDPTTYSYGLVLTALNPSPIAGQWPPAQDFQVLAMARPAV- 476
Db 1028 RQTKPAGVTPAPTSTPTPTPTPTAPTSAPTSPVTAPTPTPTPTAPTPTPTPTPT 1087
QY 477 -----VSGILKQYKKNDSAPGDNOIKPGLQLVNTGSSSDLS 514
Db 1088 PTPPTPTPTVTPPTPTPTCTPGTSGILKYLKNNETSASTGIRPFKLVNGSSSDLS 1147
QY 515 TYTVRYWFTRDGGSSTLYNCDMAAMGCGNIRASFGSVNPAPTADTYLQLSFTGT--L 572
Db 1148 RKIKRYWYVVDGDKPQSAV--CDMAQIGASNVYFNFVKLTSGVSGADYYLEVGFSSGAGQL 1206
QY 573 AAGSGTGEIQNVKSDWNSFTETNDYSY--GINTTFQDMTKTYVYVNGVLYWG----- 624
Db 1207 QPGKDTGDIQVRFKNKNDMSNYQADDMWSMQSMTNNGENAKVTLYYDGLVWGOEPGAT 1266
QY 625 ---TEDVTPPSVPTGLVYTVGVSGSSVSLAMNASTDNVGAHYNVYRNGVLYVGQPTVTSPT 681
Db 1267 PAPSTATPTPTPTATVT-----PPTPTA--T 1290
QY 682 DTGLAGTAYTYTVAAVDAAGNTSAPSTPVDCPTGPNONGVTSVODG----- 728
Db 1291 PTPPTPTPATPTPTPT-----PTTPTPKYVKRIP--PNNNPLISHKFGADPAVLVYGRYV 1344
QY 729 -----EYRQVTNEMWSSAQOCLTINTATG---ART-----VSTANBSG 763
Db 1345 MYLTNDILEYDENGWVKNDSYSKINKITIVYSSDDLVMNTDGEIEVAGPENGVAKMANFS- 1403
QY 764 GTGGAAP--TYPST-----YKCGHWGNCYTKNVMPIQISQISAVTSMSTTOVSSGA 814
Db 1404 ---WAPAVAYKINKNDREFLYFG-----NSGGGIGVITTADSPGTGPMSD----- 1444
QY 815 YDAVDIMWTNSTPTTGTQPNGTETIMWL 842
Db 1445 -PLGRPLITWSTP-----GVQGVVWL 1464
```

Search completed: August 29, 2002, 16:24:14
Job time: 349 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM,protein - protein search, using sw model

Run on: August 29, 2002, 16:38:35 ; Search time 38.44 Seconds
(without alignments)
1050.585 Million cell updates/sec

Title: US-09-917-384-6
Perfect score: 5336
Sequence: 1 MERTQSGRCNCRTRGTRM.....GFNGSYGTAAPTLSTAS 1043

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 segs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1707	30.8	872	1 GUXA_CELFI	P50401 cellulomona
2	641	11.6	438	1 GUX3_AGABI	P49075 agarius b1
3	586	10.6	1090	1 GUXB_CELFI	P50899 cellulomona
4	561.5	10.1	1045	1 GUNB_CELFI	P26225 cellulomona
5	561	10.1	462	1 GUNB_FUSOX	P46236 fusarium ox
6	548	9.9	471	1 GUX2_TIRRE	P07987 trichoderma
7	487	8.8	747	1 GUND_CELFI	P50400 cellulomona
8	449.5	8.1	1742	1 GUNA_CALSA	P22534 caldocellum
9	434	7.8	1331	1 MANB_CALSA	P22533 caldocellum
10	414	7.5	562	1 GUN1_ACICE	P25833 actinobacteriu
11	413	7.5	1853	1 CIPB_CLOCTM	P06851 clostridium
12	401	7.2	700	1 GUNB_CLOCTM	P01866 clostridium
13	388.5	6.9	1039	1 GUNB_CALSA	P29719 paenibacill
14	382	6.9	499	1 GUN3_BACSU	P10474 c endoglucan
15	381.5	6.9	499	1 GUN3_BACSU	P23549 bacillus su
16	380	6.9	145	1 YCEA_PAEIA	P29718 paenibacill
17	379	6.8	499	1 GUN1_BACSU	P07983 bacillus su
18	365.5	6.6	504	1 GUN4_THERFU	P26222 thermomono
19	365.5	6.6	504	1 GUN4_THERFU	P26222 thermomono
20	365.5	6.6	504	1 GUN4_THERFU	P26222 thermomono
21	355.5	6.1	441	1 GUN2_THERFU	P26222 thermomono
22	338.5	5.9	986	1 GUN2_CLOS	P26599 clostridium
23	325.5	5.7	699	1 CHIA_BACCI	P20533 bacillus c1
24	315	5.7	699	1 CHIA_BACCI	P20533 bacillus c1
25	309	5.6	1848	1 GUX2_CLOCT	P50900 clostridium
26	307.5	5.6	914	1 GUN2_CLOS	P26414 microspor
27	299.5	5.4	456	1 GUNA_MICBI	P27050 bacillus c1
28	289	5.2	879	1 GUN1_CLOCTM	P02934 clostridium
29	276.5	5.0	524	1 GUN1_CLOCTM	P02934 clostridium
30	273.5	4.9	962	1 GUNA_PSEPL	P10476 pseudomonas
31	261	4.7	488	1 PHB_ALCPA	P12625 alcaligenes
32	257.5	4.6	2021	1 OMPA_RICCN	P02657 rickettsia
33	253	4.6	484	1 GUX_CELFI	P07986 cellulomona

ALIGNMENTS

34	248.5	4.5	449	1 GUNA_CELFI	P07984 cellulomona
35	244	4.4	644	1 XYND_CELFI	P54865 cellulomona
36	235.5	4.2	1148	1 ICER_PSEEX	O30611 pseudomonas
37	234	4.2	237	1 GUN_ASPAC	P22659 aspergillus
38	234	4.2	597	1 CHIA_STROI	P06638 streptomyces
39	233.5	4.2	1200	1 ICEN_PSEEX	P36909 streptomyces
40	230	4.2	748	1 GUN_C_PSEPL	P27033 pseudomonas
41	228.5	4.1	1196	1 ICEN_PSEEX	O33479 pseudomonas
42	228.5	4.1	1861	1 APU_THERFU	P38336 t amylopull
43	226.5	4.1	610	1 CHIT_STRPL	P11220 streptomyces
44	226	4.1	515	1 GUND_CLOCT	P26823 clostridium
45	224.5	4.1	515	1 GUND_CLOCT	P26823 clostridium

RESULT 1
GUXA_CELFI STANDARD: PRT: 872 AA.
ID GUXA_CELFI
AC P50401;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Exoglucanase A precursor (EC 3.2.1.91) (Exocellulohydrolase A)
DE (1,4-beta-cellulohydrolase A) (CBP95).
GN CBHA.
OS Cellulomonas flm1.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1708;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 484;
RX MEDLINE=9434030; PubMed=8065260;
RA Meinke A., Gilkes N.R., Kwan E., Kilburn D.G., Warren R.A.J.,
RA Miller R.C., Jr.;
RT "Cellulohydrolase A (Cbha) from the cellulolytic bacterium
RT Cellulomonas flm1 is a beta-1,4-exocellulohydrolase analogous to
RT Trichoderma reesei CBH II".
RT Mol. Microbiol. 12:413-422(1994).
RN (2)
RP SEQUENCE OF 41-58.
RX MEDLINE=93209933; PubMed=8458833;
RA Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C., Jr., Warren R.A.J.;
RT "Cellulose-binding polypeptides from Cellulomonas flm1: endoglucanase
RT D (Cend), a family A beta-1,4-glucanase".
RT J. Bacteriol. 175:1910-1918(1993).
CC -1- FUNCTION: THIS ENZYME HYDROLYSES 1,4-BETA-D-GLUCOSIDIC LINKAGES OF
CC CELLULOSE. WEAK ACTIVITY AGAINST CARBOXYMETHYLCELLULOSE, BACTERIAL
CC MICROCARSTALLINE CELLULOSE AND BARLEY BETA-GLUCAN. HAS ALSO WEAK
CC ENDOLUCANASE ACTIVITY. HYDROLYSES GLUCOSIDIC BONDS WITH INVERSION
CC OF ANOMERIC CONFIGURATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY B (FAMILY 6 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: L25809; AAC36898.1; -
CC HSSP: P07986; IEXG.
CC Interpro: IPR001919; CBD_2.

Tue, Sep 3 13:42:07 2002

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC EMBL: I24519; AAAS0607.1; -
 CC EMBL: I24520; AAAS0608.1; -
 CC EMBL: I24507; CAAB3971.1; -
 CC HSP: P00725; IAZ6.
 CC InterPro: IPR000254; CBD_fungal.
 CC InterPro: IPR001524; Glyco_hydro_6.
 CC Pfam: PF00734; CBD_1; 1.
 CC Pfam: PF01341; Glyco_hydro_6; 1.
 CC PRINTS: PR00733; GHYDRLASE6.
 CC PRODOM: PD001821; CBD_fungal; 1.
 CC PRODOM: PD003733; Glyco_hydro_6; 1.
 CC SMART: SM00236; fCBD; 1.
 CC PROSITE: PS00562; CBD_FUNGAL; 1.
 CC PROSITE: PS00655; GLYCOSYL_HYDROL_F6_2; 1.
 CC PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
 CC Cellulose degradation: Hydrolyase; Glycosidase; Signal.
 CC SIGNAL: 1
 CC CHAIN: 21 438
 CC DOMAIN: 21 59
 CC DOMAIN: 60 87
 CC ACT_SITE: 215 215
 CC ACT_SITE: 393 393
 CC DISULFID: 39 45
 CC DISULFID: 170 229
 CC DISULFID: 360 407
 CC VARIANT: 133 133
 CC VARIANT: 152 152
 CC VARIANT: 244 244
 CC VARIANT: 248 248
 CC VARIANT: 398 398
 CC SEQUENCE: 438 AA; 46209 MW; 002C973544893794 CRC64;

Query Match 11.68; Score 641; DB 1; Length 438;
 Best Local similarity 35.08; Pred. No. 1.2e-28;
 Matches 162; Conservative 64; Mismatches 145; Indels 92; Gaps 14;

CC 14 ORGTMRRAISKRLAGVLAGAVSIAISYPLAMQHPAIAATHVDPRYAGFEFVNPYA 73
 CC 68 QGRT-PRATTSSTGSGTSGA-----GNFTYTKTWLSFPA 104
 CC 74 QEVSEANQNTATLAAMRYSTYSTAVMDRIAINGVNGPGLTYYIDALLSQOQGT 133
 CC 105 DEVAQAADADISNPSTAAASAKIPTVWFDYAKV-----PDLCGYLADARKSKQ-- 156
 CC 134 TPEVIVIVYDLPGRDCALASNGELPATPAAGLQTYETQYIDPASTILSPKYSLSLRTV 193
 CC 157 ---LVQIVYDLPGRDCALASNGELPATPAAGLQTYETQYIDPASTILSPKYSLSLRTV 210
 CC 194 IIEPDSIPNAVYMSQACATAPVYEGIEFALTKLHAIPNVYIYMDAHSGLWGPNN 253
 CC 211 VIEPDSIPNAVYMSQACATAPVYEGIEFALTKLHAIPNVYIYMDAHSGLWGPNN 269
 CC 254 ASGVGVQKVLNLSIGVNGIDGFTVNTANTPIKEPFMTATQVGGQVSEANFYQNP 313
 CC 270 LSPAQAQFAQIYRDAGSPRLRGIAIVNANFNALRMS-----SPDIT-----QGSN 316
 CC 314 DIDADVAVDLSRLVAAGPSSIGMLIDTLRNGMGCPNEPRTATPNTFVQSKID 373
 CC 317 NYDEHYEALAPMLSNAGPAH--FLVDGRSG-----VON 351
 CC 374 LROHGLWCNONGAGLGPQASPTDPNNAHLAAVYIKPGSDGSAASDPTTGKKS 433

DB 352 IRDQGMCMVKAGRGQR---TTWNGSLIDAIWVNRGEGCDGTDNSNP-----RFD 404
 CC 434 PMCDPTTSTGYVTNALPNSPIAGOMFPNPFQDOLVANNRPAY 476
 CC 405 SHCS-----LSDAHQAPAEAGTWFOAYETLVANNPFL 438

RESULT 3
 GUXB_CELFT STANDARD; PRT; 1090 AA.

AC GUXB_CELFT
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Exoglucanase B precursor (EC 3.2.1.91) (Exocellulohydrolase B)
 DE (1,4-beta-cellobiohydrolase B) (CBP120).
 GN CBH OR CENE.
 OS Cellulomonas flm1.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
 OX NCBI_TaxID=1708;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 456-461.
 RC STRAIN=ARCC 484;
 RX MEDLINE=96003998; PubMed=7575482; Miller R.C. Jr., Warren R.A.J.,
 RA Shen H., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.,
 RT "Cellulohydrolase B, a second exo-cellulohydrolase from the
 RL cellulolytic bacterium Cellulomonas flm1.";
 RN Biochem. J. 311:67-74(1995).
 RN [2]
 RP SEQUENCE OF 54-75.
 RX MEDLINE=93209933; PubMed=8458833; Miller R.C. Jr., Warren R.A.J.,
 RA Melnik A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.,
 RT "Cellulose-binding polypeptides from Cellulomonas flm1: endoglucanase
 RL D (Cend), a family A beta-1,4-glucanase.";
 RN J. Bacteriol. 175:1910-1918(1993).
 RN [3]
 RP SEQUENCE OF 54-78.
 RX MEDLINE=94197708; PubMed=8147863;
 RA Shen H., Romme P., Melnik A., Gilkes N.R., Kilburn D.G.,
 RT Warren R.A.J., Miller R.C. Jr.;
 RL "Stereochemical course of hydrolysis catalysed by Cellulomonas flm1
 RT Cend, a member of a new family of beta-1,4-glucanases.";
 RL Biochem. Biophys. Res. Commun. 199:1223-1228(1994).
 CC -1- FUNCTION: HYDROLYSE CELLULOSE TO A MIXTURE OF CELLOTETRAOSE,
 CC CELLOTRIOSE AND CELLOBIOSE, WITH ONLY A TRACE OF GLUCOSE. IT
 CC HYDROLYSED CELLULOSE TO CELLOTRIOSE AND CELLOBIOSE, AND
 CC CELLOTETRAOSE TO CELLOBIOSE, BUT IT DID NOT HYDROLYSE CELLOTETRAOSE.
 CC HAS ALSO WEAK ENDOGLUCANASE ACTIVITY. HYDROLYSES GLUCOSIDIC BONDS
 CC WITH INVERSION OF ANOMERIC CONFIGURATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellotetraose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 CC -1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS,
 CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
 CC (CBD).
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY L (FAMILY 48 OF GLYCOSYL
 CC HYDROLASES).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR Pfam: PF00553; CBD_2; 1.
 DR Pfam: PF00041; fn3; 3.
 DR Pfam: PF02011; Glyco_hydro_48; 1.
 DR PRINTS: PR00014; ENTRYPELIT.
 DR PRINTS: PR00844; GLHYDRLASE48.
 DR ProDom: PD011903; Glyco_hydro_48; 1.
 DR SMART: SM00660; FN3; 3.
 DR PROSITE: PS00561; CBD_BACTERIAL; 1.
 DR Cellulose degradation; hydrolase; Glycosidase; Repeat; Signal.
 KM SIGNAL
 FT PROPEP 34 53
 FT CHAIN 54 1090
 FT DOMAIN 54 699
 FT DOMAIN 700 785
 FT DOMAIN 794 884
 FT DOMAIN 891 978
 FT DOMAIN 899 1090
 FT ACT_SITE 513 513
 FT DISULFID 990 1089
 SQ SEQUENCE 1090 AA; 114829 MW; 046BB9D9562E399 CRC64;

Query Match 10.6%; Score 586; DB 1; Length 1090;

Best Local Similarity 24.0%; Pred. No. 4.3e-25;

Matches 299; Conservative 147; Mismatches 419; Indels 382; Gaps 62;

QY 20 MAISKRLAGVLAGAVSLAIVPLAMOHPAIAATHDNPAGATFFVNPYMAOEVOSE 79
 DB 1 MSSTRRRSAMVAATVGV-SSFLAVAGITPAIAAGAGQP---AT-----VTVP 46
 QY 80 ANQTNATLAAK--MRVSTYSTAVMMORIAINGVSGGLTTLDAALSOOGGTPPEV 137
 DB 47 MASPRAAVDEYAOEFLAQY-----DKIK--DPANG-----YFSA-----QGLPYHA 87
 QY 138 IEIVTYDLP--GRDCAALASN-----GELPATAGL-QTYEN-----QYIDPIA 178
 DB 88 VETLAEAPDYGHETTSSEASYMLMLEALYGOVQDMAPLANMPTMEKYMIPQSDQPT 147
 QY 179 SILSNPKYSSLRIVITIEEDSLPNAVTN-----MSIQACATAVPYEGGIEFALTKL 230
 DB 148 NSFYFNPNSPATYAPPEFNHSSYSPQINSGISGTDPIGELKAT-----YGNADVOMHWL 203
 QY 231 HAIPNYYIYMDAAHSG-WLGMFNNAAGYQVOEYKVLNASI-----GVNGCI 274
 DB 204 ADVDNITGCGATPGACCTIGPATGTSFINTQRGPOESVWETVPOPSCEERKYGSKCY 263
 QY 275 DGFVTNTATYPLKEPEMTATQOVGQPVES---ANFYQWNPIDEA----- 318
 DB 264 IDLFTRKASYA--KQWKYTSASDADARAFAEAYWMAN--QWATEQGRADVAATVAKAKA 319
 QY 319 -Y-----AVDLYSRV-----AAGFPSSIGMLDITLNGGGRPE----- 353
 DB 320 GYLKRYTLEFDKYFKKIGCTSPCAAGQGRHAHYLLSWYMAWGATDSSGNAMRIGSSH 379
 QY 354 -----PTGPSTATDVNTFVNQSKIDLRQHGLMCNONGAGL--- 389
 DB 380 AHFGYQNPPLAAMALSTDKLTPKSPRAKADMAAM-QRLPEY-----TWLQASNGCIAG 434
 QY 390 -----GPPQASPDDEFNAHLDAVYWKIPGES-DGTSA-----ASDPTT 428
 DB 435 ATNSMDGAVAPAGPAGTFYGMGYTEAPYVDPSPSRMTGQMAMGQVRAVELIYASGNMQ 494
 QY 429 GKKSDDMCDPTTYSYGVLTNALPNSPIAGQFPAQFDOLVANARPAVYSGGLKYOYKA- 487
 DB 495 AAK-----IIDKWPVWVAVIISTDGASWKPSEIKWTGKPD 530
 QY 488 --NDASPGDNQIKPGLQ-LVNTGSSVYDLSVTYVR---YMFTRDGGSSR-----LYNCD 536
 DB 531 TMAAAPTGN---PGLITVEVTSYGODVGYAADTARALLFYAAKSGDTARAKALLDII 587
 QY 537 WAMAGCGNIRASGYNPATPTADTYLQISFT-----GGTLAAGSGTEIGION-RVNSKDM 590
 DB 588 WA-----NNQDPLGV--SAVETRGDYKKRPDDTYVANGSDIYIPSGWTGIMPNGDVAKPEV 640

QY 591 SNETETNDYSTGTNTTFDDMKIVYVNG-----VLYWG----- 624
 DB 641 SFLDIRSEFYKKDPN-----WSKVQTFELDGAEPQFRYHRFMAQTAVAGALADYARLFDG 695
 QY 625 --TEDVTPSPVPTGLVTVGVSSGSSVSLAMNASTDVGVAHVYVYVNGVLPYTSPTD 682
 DB 696 TTPPTPTATPTVPTGLQAGVYVSTENATISWTASTDTRVYTGIDVYKAGATVATTSTFD 755
 QY 683 TGLAGTAVTVYVAADVADAGNTSAPS-----TPYDCPTGPNQNGVTSVQDEYHVQ 733
 DB 756 TGLTASTAVYAVYVRAFDAGNVASPSALVTYTKATPSD-TTAPSPVATIS----- 805
 QY 734 TNEHNSAQOCLITITATAGATVSTANFSGGTGGAPATYPSLIYKCCMHGNCCTTKAVGMP 793
 DB 806 -----SSS-----TANSVITGMSASTDN-AGGSG-LAGY-DVYRGA-----T 839
 QY 794 QISQISAVTSMSTQVSSGAY-----DVYDI-----WTNSPTTGTG--P 833
 DB 840 RYAQT-TALTFTDTGLTSTAYETVRADVAGNVASSTAVSTTKSDTTPDTTAPSV 898
 QY 834 NCTEIMIMLNSRGVQPRGSOATATGTVVAGHTVNVMOGQTSWKIISYVLAPGATSISNL 893
 DB 899 AGIAMA-----TVTEISYA-LHW-----ASTDTGGS 924
 QY 894 DKA--TFADAAARSLNTSYL--LDVEAGFE-----TWGGGLGNSNSVSYS-- 941
 DB 925 GLKGYDVYRGATRVGTSYASTYDGLTAATAYQYTVRATDAGNVASASALSTYTKTP 984
 QY 942 -GGVACRAITYYVSDMWSGFTATVTVNTGSRATGVMVMSFGNGQVYTNVNTNALQS 1000
 DB 985 QTGSGCSVAIYANSS-WNSGFTASVRITNGITTINGMSIGFLTGQKQVQGSMTWTQS 1043
 QY 1001 GASVTATNLYSNNVIOPOGSTTFGNGSYSGTNAAPT-----LSCT 1041
 DB 1044 GSTVTATNAPMNGTLAPGTVDVGFNGSHGTGNPAPSTLNGASCT 1090

RESULT 4
 GUNB_CELFI STANDARD; PRT; 1045 AA.
 AC P26225;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase B)
 GN CENB.
 OS Cellulomonas fimi.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Micrococciaceae; Cellulomonadaceae; Cellulomonas.
 OX NCBI_TaxID=1708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91100298; PubMed=1987122;
 RA Melnik A., Braun C., Gilkes N.R., Kilburn D.G., Miller R.C. Jr.,
 Warren R.A.J.;
 RT "Unusual sequence organization in CenB, an inverting endoglucanase
 from Cellulomonas fimi.";
 RT J. Bacteriol. 173:308-314(1991).
 RN [2]
 RP MEDLINE=92041609; PubMed=1938913;
 RX Melnik A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
 RT "Multiple domains in endoglucanase B (CenB) from Cellulomonas fimi:
 RT functions and relatedness to domains in other polypeptides.";
 RT J. Bacteriol. 173:7126-7135(1991).
 CC (1) FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
 CC (1) ENDOGUCCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
 CC (2) EXOCELLULOIOHYDROLASES THAT CUT THE DISACCHARIDE CELLULOSE
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER

SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.

-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.

-1- MISCELLANEOUS: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL SITE FOR PROTEOLYSIS.

-1- MISCELLANEOUS: MAY CONTAIN A SECOND CYPD IN THE CATALYTIC DOMAIN.

-1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.

-1- SIMILARITY: BELONGS TO CELLULOSE FAMILY E (FAMILY 9 OF GLYCOSYL HYDROLASES).

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-slb.ch/announce/> or send an email to license@isb-slb.ch).

CC EMBL: M64644; AAA23086.1; -

CC PIR: A39199; A39199.

DR HSSP: P26221; ITR4.

DR InterPro: IPR001919; CBD_2.

DR InterPro: IPR001956; CBD_3.

DR InterPro: IPR003961; FN_III.

DR InterPro: IPR001701; Glyco_hydro_9.

DR Pfam: PF00553; CBD_2; 1.

DR Pfam: PF00942; CBD_3; 1.

DR Pfam: PF00041; fn3; 3.

DR Pfam: PF00759; Glyco_hydro_9; 1.

DR SMART: SM00060; CBD_BACTERIAL; 1.

DR PROSITE: PS00561; CBD_BACTERIAL; 1.

DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.

DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_2; 1.

DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.

KW Cellulose degradation: Hydrolase; Glycosidase; Repeat; Signal.

KW POTENTIAL.

FT SIGNAL 1 33

FT CHAIN 34 1045

FT DOMAIN 34 643

FT DOMAIN 644 650

FT DOMAIN 651 733

FT DOMAIN 734 748

FT DOMAIN 749 830

FT DOMAIN 831 846

FT DOMAIN 847 930

FT DOMAIN 931 944

FT DOMAIN 945 1045

FT ACT_SITE 410 410

FT ACT_SITE 449 449

FT ACT_SITE 458 458

FT DISULFID 946 1044

FT SEQUENCE 1045 AA; 108990 MW; AC2F7B84E4E3C4F0 CRC64;

Query Match 10.1%; Score 561.5; DB 1; Length 1045;

Best Local Similarity 26.7%; Pred. No. 9,3e-24;

Matches 265; Conservative 105; Mismatches 382; Indels 241; Gaps 47;

QY 146 PGHD-----CAALASN-----GELPATAAGL-----QYENQYIDPIKSIISNP 184

DB 181 PGSDVAETRAALASAIYILKGDDEPAYATLVSNAKOLYTPADYRGATISQCVRAASAY 240

QY 185 K-YSSLRTVITTEPDSLPNNVNMISQACATVAPYEOGIEYALKLHAIPNYIYMDAA 243

DB 241 KSMGSGYODELWGAAYWLYKATGDATYLAKEA-EYDKLGTENOST-----TRS 287

QY 244 HSGWLGMPNNASG-----YQEVQVQLNA-SICVNGIDGFTNTANTPLKE 289

DB 288 YKMTIANDNKOFYTYALLAMETGKOKYVDANWMLDYTWVGNG-----QKVPYSP--- 338

QY 290 PPMATQOVGGQPY-----SSAN-----FYQMNPD-----IDEADYAV- 322

DB 339 -----GGQAVLDKMGALRYAANTSFVALYVSDMMTDATKARAYHFGVAKQIYVAG 389

QY 323 -DLYSRLVNAAGFPSSIGMLIDTLRNGWGGPNEPTGISTADVTNFNQSKIDLRQRGLM 381

DB 390 DNPSSSYVGE-----GANPPTAPHHRTAHGSMWLDSTTPPAQSHRYLV 433

QY 382 CNONGAGLCGPQASPTDFPNAHIDA-----YWIKPPGSDSTSAASPTTGKSKDPCMP 438

DB 434 ----GALVGGP---GSPND---ATDSRODY-----ANEAATDYNAG----- 466

QY 439 TYTTSYGLVTLNLPNSPIAGWEPNO-----FDQVANARPAVYSGGLKQYKNNDSAP 492

DB 467 -FTSALARLVEEGCGRPFLPPEQPDGDLFVAMLNQPSGTFTEVKMIRNQSAPF 525

QY 493 GDNQIKPGLQLVITGSSVDLSVTVRYWYFRDG-GSTLYVNCDMAAGCGNIRASFGS 551

DB 526 ARS-----LNNAKRYFTTIDGFASQVTLNAYTSE--CG--AOSGK 563

QY 552 VNPATPADYTLQISFTGTLAAGST--GEIQNRV-NKSPMSNFTENDVSYG--TNT 605

DB 564 GVSAGGTIG-YVELSCVGODIHFGGOSQHRRETOFLTGDPAGN--PANDSYGLVOT 619

QY 606 TPODMKVTVYVNGVLYWG-----TEDVTPSPVPTGLVTVGVSSSVSLANASTD-NV 658

DB 620 ALAKASATITLYDGSSTLVWGKEPTGTTDTTPPTPGPATGVTGASISMAASTDACS 679

QY 659 GVAHNVYR-----NGVLVGOPTVTSFTDTGLAGATATTVYAAVDAAGNTSAPSTPVDCT 714

DB 680 GVAGIELRVGGTGTQTLVGTTLAAAYILRDLTPGTAVSYVKKADVAGNVSASAVT-- 737

QY 715 PCPNQGVTSVODGRTYRQTNEMNMSAQOCLITWATGAMTVSTANPSGCTGAPATYDS 774

DB 738 -----FTTDTTGTETPTTP-GTPVASAVTSGATLAMPST-----GDPAY--S 779

QY 775 IYKCGHWGNCITKKNVGMPIQISQISAVTSWSTVOVSSGADVAYDINTSTP--TTTG 831

DB 780 GYVLVAVGCTTTTVAQTVPTVLTSLGLPSTAYVAVRANVAGDVASAPVTTTAA 839

QY 832 QPNGTEIMTLMSKRGVOVF--GSQTATGVYVAGTMMVMVMOGQOTSW--KIISVLRPGA 867

DB 840 PVDVT-----VAFTVPTPYASVAVATTGAIL-TWNASDSSGSGLAGREVLRVS 887

QY 888 TSISNLDLKAIFDAARSLNTSDVLLDVEAGFEIWOGGCGIGNSFSVTSGGV--- 944

DB 888 GTTQTLVASTTAVLALGLTPATAYVYRAK---DAGNVASVSPVFTTLPYIST 943

QY 945 -ACRATTVVNSDMGSGTATVYVNTGSRATNGVAVMSFGNGQTVYVNTALTOGAS 1003

DB 944 PSCTVVTVSTNS-NWVGSTGSKITNTGTTPLT-WTIGFAFPGGQVOTQGSATWSQGTGT 1001

QY 1004 VTATNLYSNVYIOPGOSTTEFGNFSYSTNAAP 1036

DB 1002 VTATGLSMNATLQPGQSTIDIGFNGSHPTNTNP 1034

RESULT 5

GUNB_FUSOX STANDARD: PRT: 462 AA.

AC P46236:

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE Putative endoglucanase type B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).

OS Fusarium oxysporum.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreales; mitosporic Hypocreales; Fusarium.

OC NCBI_TaxID=5507;

RN (1)

RP SEQUENCE FROM N.A.

RX MEDLINE=95047531; PubMed=7959045;

RA Sheppard P.O., Grant F.J., Oort P.J., Sprecher C.A., Foster D.C., Hagen F.S., Upshall A., McKnight G.L., O'Hara P.J.;

RT The use of conserved cellulase family-specific sequences to clone cellulase homologue cDNAs from Fusarium oxysporum.;

RL Gene 150:163-167(1994).

CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- SIMILARITY: CONTRAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: L29377; AAA65585.1; -.
 CC HSSP: P07987; ICB2.
 CC InterPro: IPR000254; CBD_fungal.
 CC InterPro: IPR001524; Glyco_hydro_6.
 CC Pfam: PF00734; CBD.1; 1.
 CC Pfam: PF01341; Glyco_hydro_6; 1.
 CC PRINTS: PR00733; GLYDRLASE6.
 CC ProDom: PD001821; CBD_fungal; 1.
 CC ProDom: PD003733; Glyco_hydro_6; 1.
 CC SMART: SM00236; ICB2; 1.
 CC PROSITE: PS00562; CBD_FUNGAL; 1.
 CC PROSITE: PS00555; GLYCOSYL_HYDROL_F6_2; 1.
 CC PROSITE: PS00556; GLYCOSYL_HYDROL_F6_2; 1.
 CC Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
 CC SIGNAL 1 16
 CC FT CHAIN 17 462
 CC FT DOMAIN 17 65
 CC FT DOMAIN 66 99
 CC FT DOMAIN 100 462
 CC FT ACT_SITE 190 190
 CC FT ACT_SITE 236 236
 CC FT ACT_SITE 416 416
 CC FT DISULFID 33 50
 CC FT DISULFID 44 60
 CC FT DISULFID 191 250
 CC FT DISULFID 383 430
 CC FT CARBOHYD 37 37
 CC FT CARBOHYD 223 223
 CC FT CARBOHYD 272 272
 CC FT CARBOHYD 317 317
 CC SEQUENCE 462 AA; 49207 MW; E25B2F5B828B637F CRC64;

Query Match 10.1%; Score 561; DB 1; Length 462;
 Best Local Similarity 32.5%; Pred. No. 3.4e-24;
 Matches 150; Conservative 71; Mismatches 167; Indels 74; Gaps 16;

QY 17 TTRMPAISRRLRAGVLAGAVSTAASIVPLAMQHPATAATHVDPYAGATFVNPVWAOEV 76
 DB 70 TAAGSSSTATRTATGSGSTTGAAGSVTSA--PPAAS--DNPYAGVDLMANNYRSEY 123
 QY 77 GSEAA---NQTNATLAAKKRVSTVSTAVMDRIAMINOVNGRPGLTYYLDAALSOOGFT 133
 DB 124 MNLAIVPKLSGAKATAAKAVDPSPFQ--WMDYDHISLME---DTLADIRKANKAG- 174
 QY 134 TREVEIIVYIDLPGRDCALASNGELPATRAGIQETGYIDPIIASILSNPKYSLRIYV 193
 DB 175 GYVAGGFVVYDLPNRDCAAAANSNGESLDKDKANKYKA-YIAKIKILQN--YSDTKVIL 231
 QY 194 IIEPDSLPAVATNMSTIOACATVAPYEEGIEVALKRLHAIPVYIYMDAAHSGMLGMPNN 253
 DB 232 VIEPDSLAVNLNLYNDKCAKAEAKELTYAIKELN-LPNVSMYLDAGHGMLGMPAN 290
 QY 254 ASGYVQEVQKVLNASTGVNGIDGFTVNTANTYTPLEKPEFTATQOVGGQVEANFQOMP 313
 DB 291 IEPAAALVIAQIYKKDAKPRVRGVLVNVSNY-----GWLKSTRPQDYTESNP 337
 QY 314 DIDEADYAVDLISRLVAAGFPSSIGMLIDTLRNGMGNGPEPTGATVNTFVNGSKID 373

DB 338 NYDEORY-INAFAPLLAOGKSNVKRTYDGRSG-----KQPIG----- 375
 QY 374 LNRHRLMCNQGAGIGQPPQASPTDFPAHLDAYVWIKRPESDQTSASPTGKSD 433
 DB 376 -KQAGDCMNAKGTGFLRPSTWTGD--ALDAFVWVRPGESDQTS-----DTSARVD 427
 QY 434 PMDDPYTTSYGVLTALNSPIAGWFPAPQADQVANA RPA 475
 DB 428 YHCG-----LDDALKPAPAGTGWFAYPEQLDNNAPS 460

RESULT 6
 GUX2_TRIRE STANDARD; PRT; 471 AA.
 AC P07987;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Exoglucanase II precursor (EC 3.2.1.91) (Exocellulohydrolase II)
 DE (CBHI) (1,4-beta-cellulohydrolase).
 GN CBH2.
 OS Trichoderma reesei (Hypocrea jecorina).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; Hypocreaceae; Hypocrea.
 OX NCBI_TaxID=51453;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VTT-D-80133;
 RX MEDLINE=87248061; PubMed=3596237;
 RA Teeri T.T., Lehtovaara P., Kauppinen S., Salovuori I., Knowles J.,
 RT sequence and expression of cellulohydrolase II.";
 RL Gene 51:43-52(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=QM9414 / RUT C-30;
 RX Chen C.M., Gritzal M., Stafford D.W.;
 RT "Nucleotide sequence and deduced primary structure of
 RL cellulohydrolase II from Trichoderma reesei.";
 RN Biotechnology 5:274-278(1987).
 RN [3]
 RP SEQUENCE OF 25-44.
 RA Paegerstam L.G., Petersson L.G.;
 RT "The 1,4-beta-glucan cellulohydrolases of Trichoderma reesei QM
 RT 9414.";
 RL FEBS Lett. 119:97-100(1980).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=90333255; PubMed=2377893;
 RA Rouvinen J., Bergfors T., Teeri T.T., Knowles J.K.C., Jones T.A.;
 RT "Three-dimensional structure of cellulohydrolase II from Trichoderma
 RT reesei.";
 RL Science 249:380-386(1990).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=97029636; PubMed=8873646;
 RA Koivula A., Reinikainen T., Ruohonen L., Valkeajarvi A.,
 RA Claessens M., Telemann O., Kleywegt G.J., Sardenings M., Rouvinen J.,
 RA Jones T.A., Teeri T.T.;
 RT "The active site of Trichoderma reesei cellulohydrolase II: the role
 RT of tyrosine 169.";
 RL Protein Eng. 9:691-699(1996).
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
 CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
 CC (2) EXOCELLULOHYDROLASES THAT CUT THE DISACCHARIDE CELLULOSE
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellotetraose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- MISCELLANEOUS: J. RESEI PRODUCES TWO DIFFERENT
 CC EXOCELLULOBIODHROLYASES. THEY ARE UNIQUE IN THAT THEY CAN HYDROLYSE
 CC CRYSTALLINE CELLULOSE IN THE ABSENCE OF ENDOGLUCANASES.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL
 CC HYDROLASES).
 CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M16190; AAA34210.1; -;
 CC EMBL: M55080; AAA72922.1; -;
 CC PIR: A26472; A26472.
 CC PIR: A26160; A26160.
 CC PDB: 3CBH; 15-JAN-91.
 CC PDB: 1CB2; 19-MAR-99.
 CC InterPro: IPR000254; CBD_fungal.
 CC InterPro: IPR001524; Glyco_hydro-6.
 CC Pfam: PF00734; CBD_1; 1.
 CC Pfam: PF01341; Glyco_hydro-6; 1.
 CC PRINTS: PR00733; GHYDRASE6.
 CC ProDom: PD001821; CBD_fungal; 1.
 CC ProDom: PD003733; Glyco_hydro-6; 1.
 CC SMART: SM00236; ICB; 1.
 CC PROSITE: PS00562; CBD_FUNGAL; 1.
 CC PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
 CC PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
 CC Cellulose degradation: Hydrolase; Glycoprotein; Signal;
 CC 3D-structure. 1 24
 CC SIGNAL. 1 471 EXOGLUCANASE II.
 CC CHAIN. 25 471 CELLULOSE-BINDING (BY SIMILARITY).
 CC DOMAIN. 25 65 LINKER.
 CC DOMAIN. 66 106 CATALYTIC.
 CC DOMAIN. 107 471
 CC ACT_SITE. 199 199 PROTON DONOR.
 CC ACT_SITE. 245 245 NUCLEOPHILE.
 CC ACT_SITE. 425 425 BY SIMILARITY.
 CC DISULFID. 34 51
 CC DISULFID. 45 61
 CC DISULFID. 200 259
 CC DISULFID. 392 439
 CC DISULFID. 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD. 111 111 O-LINKED (MAN. . .).
 CC CARBOHYD. 121 121 O-LINKED (MAN. . .).
 CC CARBOHYD. 130 130 O-LINKED (MAN. . .).
 CC CARBOHYD. 133 133 O-LINKED (MAN. . .).
 CC CARBOHYD. 134 134 O-LINKED (MAN. . .).
 CC CARBOHYD. 139 139 N-LINKED (GLCNAC. . .).
 CC CARBOHYD. 313 313 N-LINKED (GLCNAC. . .).
 CC CARBOHYD. 334 334 D->A: 20% OF WILD-TYPE ACTIVITY.
 CC CARBOHYD. 199 199 D->A: NO MEASURABLE ACTIVITY.
 CC MUTAGEN. 245 245 P -> R (IN REF. 2).
 CC MUTAGEN. 359 359 P -> A (IN REF. 2).
 CC CONFLICT. 449 449
 CC CONFLICT. 449 449
 CC SEQUENCE. 471 AA; 49653 MW; CA711BC35B1BD88 CRC64;

Query Match 9.9%; Score 548; DB 1; Length 471;
 Best Local Similarity 31.2%; Pred. No. 1.8e-23;
 Matches 156; Conservative 60; Mismatches 176; Indels 108; Gaps 17;

QY 7 GGRNCRQR-----GTRMPAISKRLAGVAGAVSTIASTVP-----LAMQ 48
 DB 47 SGTCTVSNDRYSOCLPGAASSSSSTRRASTSR-----VSPTSRSSSATPPPGSTTR 101
 QY 49 HPAI---AAHVDPYAGATFFVNPYAOVSEANQNTAAAKRVSTYSTAVAMD 105
 DB 102 VPPVSGSATYSGNPFVGVTPWMAAVYASVSLAIPSLTGAMATMAAAVAKVPSTWMD 161

QY 106 -----RIAALNGVNGRGILTYLDAALSOOGGTTPEVIEIYIIDLPGRDCAALAS 155
 DB 162 TLDRPLMEQTLADIRTKARKNG--NNG-----GFVYVDLPDRCCALAS 205
 QY 156 NGELPAPTAAGLQYETQYIDPIASILSNPKYSSLRITVTITPDSLPNAVTKMSIQACATA 215
 DB 206 NGEYSTADGVAKRK-NYIDTIRQIV--VEYSIDIRTLVIEDPSILANLVTLGPKKANA 262
 QY 216 VPEYEGIEYALTKLAIRNYIYMDAHSGLMGWPNNASGYVEQVQVNLASIGVNGID 275
 DB 263 QSAVLEICINAYVQLN-LPVNAVTLDRGAGHLGWPANODPAQGLFANVYKNASSPALRL 321
 QY 276 GFVNTFANYPLKEPMTATQVGGQPVESANFYQGNPDIDADYAVDLXSLVAAGPPS 335
 DB 322 GLATVNAVNY-----GMNITSPSYTQGNNAVYNEKLYTHAIGPLANHGW-S 367
 QY 336 SIGMLIDTLRNGMGCPNEPPTATDVNFVQSKIDLROHRLGMCNONGALGQPPQA 395
 DB 368 NAFPTDQGRSG---KQPTG-----QQQGMDCMNVIGTGFIRPSA 405
 QY 396 SPTDPNNAHLDAVYKIRPGESDGTSAASDPTTGKSDKPCDPTTYSYGLTNALPNSP 455
 DB 406 NTGD---SLDSFVWVRKGGEGCDTSDSAP---RDSHC-----ALPDALQAPAP 449
 QY 456 IAGQWPPAQFDQLVANARPA 475
 DB 450 QAGAWFOAYFYVOLLTNAPS 469

RESULT 7
 GUND_CELFT STANDARD; PRT: 747 AA.
 ID GUND_CELFT
 AC P50400;
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 01-OCT-1996 (rel. 34, Last annotation update)
 DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (cellulase).
 GN CEND.
 OS Cellulomonas fimi.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
 OX NCBI_TaxID:1708;
 RN [1]
 RP SEQUENCE FROM N.A. Pubmed=8458833;
 RA Weinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.;
 "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
 D (Cend), a family A beta-1,4-glucanase";
 J. Bacteriol. 175:1910-1918(1993).
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- PATHWAY: CELLULOSE DEGRADATION.
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
 CC (CBD).
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L02544; AAA3089.1; -;
 CC HSP: P07986; IEXG.
 CC InterPro: IPR001919; CBD_2.
 CC InterPro: IPR003961; FN_III.
 CC InterPro: IPR003962; FNIII_repeat.

DR InterPro: IPR001956; CBD_3.
 DR InterPro: IPR001547; Glyco_hydro_F5.
 DR Pfam: PF00942; CBD_3; 2.
 DR Pfam: PF00150; cellulase; 1.
 DR PRODOM: PD001947; CBD_3; 2.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 KM Hydrolase; Glycosidase; Cellulose degradation; Signal:
 KM Multifunctional enzyme.
 FT SIGNAL 1 41
 FT CHAIN 42 1331
 FT DOMAIN 42 325
 FT DOMAIN 326 361
 FT DOMAIN 362 518
 FT DOMAIN 519 564
 FT DOMAIN 565 720
 FT DOMAIN 721 780
 FT DOMAIN 781 1331
 FT ACT_SITE 162 162
 FT ACT_SITE 257 257
 FT CONFLICT 338 338
 FT CONFLICT 340 346
 SO SEQUENCE 1331 AA; 146892 MM; PFCASJ1B8DBF0E0 CRC64;
 POTENTIAL.
 BETA-MANNANASE/ENDOGALACTANASE A.
 CATALYTIC (MANNANASE ACTIVITY).
 PRO/SER/THR-RICH (PT BOX).
 SUBSTRATE-BINDING (POTENTIAL).
 PRO/SER/THR-RICH (PT BOX).
 SUBSTRATE-BINDING (POTENTIAL).
 PRO/SER/THR-RICH (PT BOX).
 CATALYTIC (ENDOGALACTANASE ACTIVITY).
 PROTON DONOR (BY SIMILARITY).
 NUCLEOPHILE (BY SIMILARITY).
 T -> P (IN REF. 2).
 TPPTPT -> ROHQHRO (IN REF. 2).

Query Match 7.8%; Score 434; DB 1; Length 1331;
 Best Local Similarity 23.4%; Pred. No. 1,4e-16;

Matches 215; Conservative 126; Mismatches 349; Indels 230; Gaps 41;

QY 252 NNAGSYGVOEOKVLNAGISGVNGID-GFYVTNTANTYPLKEPPTATQ---QVGGQPVESAN 307
 DB 377 NSTYTIIPMLKVVNS--GSSSIDLSRVYIRWYTVDEBQAQSAISDMAQICASNV-TFK 433
 QY 308 FYQWMPDIDEDYANDLSRLVAAGFPSSIGLIDTLRNGMGPEHPGPSTATVNTVFV 367
 DB 434 FVKLISSVSAGADYLEI-----GFKSGAGQL-----QPKDYG----- 466
 QY 368 NOSKIDLRQHRGLMCNONGAGIGPPQASPTDFPNAHLDAY-----VWIKPPGESDGTSA 422
 DB 467 ---ELQMFKNMDMKNYNOGNDMSIQMTSGENKATYATIDGLVWGQEP---SGATP 520
 QY 423 ASDPTTGKSDPMDPVTYTSYGLVTLNALPNSPIAGQWFPAPFDLVANARAVVSGGLK 482
 DB 521 AAPATATPTPTVTPVTPVTPVTPVTPVTPVTPVTPVTPVTPVTPVTPVTPVTPVTPVTP 571
 QY 483 VOYKNDASAPGNOIKPGLQLVNTGSSSVDSVTYVRYWTFPDGSSSTILYVCDMAAGC 542
 DB 572 VLYAKKENTSTNTIRPMLKVVNSGSSSIDLSRVYIRWYTVDEBQAQSAIS-DMAQIGA 630
 QY 543 GNIRAFSGSVNPAPTATYLOLSFTGT--LAAGSTGEIONRYNKSQMSNFTENDYS 600
 DB 631 SNVTFKFKLISSVSAGADYLEIGFKSGAGQLQPKDGTGEIQRNKSQMSNFTENDYS 690
 QY 601 Y-GTNTFQDMTKTVYVNGVLWGTEDEVTPPSVPTGLVTVGSSSVSLANASTDNGV 659
 DB 691 WQSMTSYGENKATYATIDGLVWGQEP---PSGTPSPSTPTVTPVTPVTPVTPVTPVTPVTP 743
 QY 660 VAHVYVYVNGVLVQGPVTSF---TDYGLACT---ATYTYAAADAGNTSAPSTPDC 713
 DB 744 TPTPTVTPVTPVTPVTPVTPVTPVTPVTPVTPVTPVTPVTPVTPVTPVTPVTPVTPVTP 801
 QY 714 TP--GPNQ-----NGVTSVQ--DGEYRVQTEVNSSAQO--CLTI----- 747
 DB 802 SPYITGANODIEGVYHSARRLGANLTGYNMFNNSNAGNDMYHSSDDYLCMSKIGISGD 861
 QY 748 -----NTATGAMVSTANF---SGGTGAPATYPSIYKCHGNCNTT 786
 DB 862 AKVPAVAVSKFHEYSLKNNAYSATVLOAGYVSKDNYGVSENEFAPS---NRNAEYVF 917
 QY 787 KNVGAPIC-----ISQISAVTSMSSTVOVSSGAVDAVDITNSTP 827
 DB 918 KK-DAPLSLNDPNDNFYVMDFFINYLKYGMA---SPGKIKGYIIDNEPDLMASTHP 973
 QY 828 TTTGGPNG-----TEIMIWLSRSGVQPFSGQTATGV--TYAGHTWVWQO 872

DB 974 RI--HPNVTCKELLEKSVELAKVITKIDPPSAEYGVGYSTGCMGYSLQDAPDMQVKG 1031
 QY 873 QTSKTIISVLP-----CATSISNLDL-----KAIPA- 900
 DB 1032 H-RW-FISWYLEQMKKASDSFGKRLDLVDLHWPEARGNIRVCFDENDTSKREVTAR 1089
 QY 901 -----DAAKRSINT-----SDYL--LDVEACFEIYWGQGGIGNSNF 935
 DB 1090 MQAPPTLMDPYKTSVVKQIDTAGENSWINQWPSDLPPIIPNVKADIEKRYYPOTKLAISEF 1149
 QY 936 SV---SVTSGVACATYVYVNSDMGSGFATVTVNTGSRATNGWTVMASFGNQTVNY 992
 DB 1150 DYGGNNHISGGIATLADVLCIFKRYGVNFAA--RMGDSGYSYAAAYNITLNYDGG--SKY 1205
 QY 993 WNTALTQSGASYATATNLXSN 1012
 DB 1206 GNTVNSAMTSVDENNPVYAS 1225
 RESULT 10
 GUNL_ACICE STANDARD: PRT; 562 AA.
 AC P54583;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Endoglucanase E1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E1)
 DE (Cellulase E1) (Endocellulase E1).
 OS Acidothermus cellulolyticus.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Frankineae; Acidothermaceae; Acidothermus.
 NX NCBI_TaxID=28049;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 43068 / 11B;
 RA Layton R.A., Himmel M.E., Thomas S.R.;
 RL Submitted (AUG-1995) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RX X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 42-398.
 RP MEDLINE=96346058; PubMed=8718854;
 RA Sakon J., Adey W.S., Himmel M.E., Thomas S.R., Kaplins P.A.;
 RT "Crystal structure of thermostable family 5 endocellulase E1 from
 RT Acidothermus cellulolyticus in complex with cellobiose."
 RL Biochemistry 35:10648-10660(1996).
 CC -1- FUNCTION: THERMOSTABLE ENZYME WITH AN OPTIMAL TEMPERATURE OF 81
 CC DEGREES CELSIUS. HAS A VERY HIGH SPECIFIC ACTIVITY ON
 CC CARBOXYMETHYLCELULOSE.
 CC -1- CATALYTIC ACTIVITY: Endohydrolays of 1,4-beta-D-glucosidic
 CC linkages in cellulose. Endoglucanase FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U33212; AA75477.1; -.
 CC PDB: 1EC6; 1A-0CT-96.
 DR InterPro: IPR001919; CBD_2.
 DR InterPro: IPR001547; Glyco_hydro_F5.
 DR Pfam: PF00553; CBD_2; 1.
 DR Pfam: PF00150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 KM Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.
 FT SIGNAL 1 41
 FT CHAIN 42 562
 FT DOMAIN 42 400
 FT DOMAIN 401 461
 PRO/SER/THR-RICH (LINKER).

DOMAIN 462 562 CELLULOSE-BINDING (BY SIMILARITY).
 ACT_SITE 203 203 PROTON DONOR.
 ACT_SITE 323 323 NUCLEOPHILE.
 DISULFID 75 161
 DISULFID 209 212
 SEQUENCE 562 AA: 60747 MW: 848625640635041 CRC64:

Query Match 7.5%; Score 414; DB 1; Length 562;
 Best Local Similarity 22.8%; Pred. No. 6e-16;
 Matches 196; Conservative 80; Mismatches 220; Indels 364; Gaps 36;

227 LFTLHAIPNYIYMDAHSGLMGPNNASGYOEVOKVLNASIGVNGIDGFTNTAN--- 283
 24 LAVALMALANLAVPRPARAAGGYW--HTSG--REILDANNVPRILGIMWFGEFCNYYV 79
 284 -----YPLKEPPMTATQVGVGPESANFVOMNPDIADAVDL 324
 80 HGLMSRDYRSMLDQIKSLGYNTIRLPSDDILKPGTMP-NSINFYOMNDL-QGLTSLQV 137
 325 YSRVYAGFPSSIGMLIDLRLNGMGPNBPTGPSTADVTENVQSKIDLROHRLMCNQ 384
 138 MDKIVA--YAGQIGL-----RIILDRHR----- 158
 385 NGAGLGGPPQASPTDFPNALIDAYWIKPPGESDGTSAADPTTGKSDPMCDPTTTSY 444
 159 -----PDCSCGSAIMW-----TSVSEAT----- 177
 445 GVLINALPNSPIAGOWPFAOFDOLVANAAPVYSGGLKVOYKNNDSAPGDNQIKPGIQV 504
 178 -----W-----ISDQLALAQ-----RKGNPTV-----VGFDLH 201
 505 NTGSSSVDLSTVTVRYWTFRDGSSSTLVYNCNW--AAMCGNIFASGVSVP----- 554
 202 NEPHDPA-----CMCGCDP-----SIDWRLAARAGN--AYLSVAPNLLIEVEG 243
 555 -ATPPADRYLQSTFGTLAAGSGTGEIQNRKNSDMNFTETDYSYGTTFQDMTKV 613
 244 VQSYNGDST---WNGNLOGAGQYPVYLVNPNRLVYS---ANDYA-----T 283
 614 TVYVNGVLWCTEDVTSPSPTGLVYVGVSGSSVSLANNASTDNGVANHVRNGVLVG 673
 284 SYVPO---TWFSDPTEPPNMP-----GIMNKNMGYLENQNAPVWLEFGT----- 326
 674 QPVTSTFD-TGLAGTATYTVAAVDAAGNTSAPSTPVDCTPGPNQGVTSVODGEYRV 732
 327 --FLQSTTDQWTKLVOLRPPA-----QYGADSFQ----- 356
 733 OTNEMNSAQQCLTINTATGAMTVSTAN-FSGGTGAPATYTSYKGMGCKTKNNG- 790
 357 -----WTFMSNPNDSGDTGG-----ILKD-DMOTVTVVKKDY 387
 791 -MPLQIS---QIGSAVTSMTQVVS---SGAYDAVDIWNTPPTTGCPNGCTEIMWLN 843
 388 LAPKKSIFDPVAGASASPSQSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 438
 844 SRGSGPFGSOFATGVTVAGHTNVMWQGOQSMKIIISYLTGCAISISNLDLKAIFADA 903
 439 -----PLIPLTAPPTPTASP-----TPSPPTAS----- 461
 904 ARGSLMTSDYLIDVEAGFELMGGGGLGNSFSVSYTSGVACRATYVNSDMGSGFTAT 963
 462 -----GARTASAGVNSDMGNGFTYV 482
 964 VTVTNGSRATNGVTAMSEFGNOTVTNYNTALTQSGASVATNLKSNVITOPGOSTTP 1023
 483 VAVTNGSGVATKWTVTSEFGNOTITNSMNAVATQNGOSVTAHMSYNNVITOPGONTTF 542
 1024 GFNGSYGCTNAAPTISCTAS 1043
 543 GFQASYSCTGSMNAPTVACAAS 562

RESULT 11
 CIPR_CLOTM STANDARD: PRT: 1853 AA.
 ID CIPR_CLOTM 006851;
 AC 01-JUN-1994 (Rel. 29, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Cellulosomal scaffolding protein A precursor (Cellulosomal glycoprotein S1/Sr) (Cellulose integrating protein A) (Cohesin).
 GN CIPR.
 OS Clostridium thermocellum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 CC Clostridium.
 CX NCBI_TaxID=1515;
 [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-40 AND 42-43.
 RC STRAIN=ATCC 27405;
 RX MEDLINE=93209931; PubMed=8458832;
 RA Gerngross U.T., Romaniec M.P.M., Kobayashi T., Huskisson N.S., Demain A.L.;
 RT "Sequencing of a Clostridium thermocellum gene (cipA) encoding the cellulosomal S1-protein reveals an unusual degree of internal homology";
 RL Mol. Microbiol. 8:325-334(1993).
 [2]
 RP SEQUENCE OF 1820-1853 FROM N.A.
 RX MEDLINE=93209931; PubMed=8458832;
 RA Fujino T., Beguin P., Aubert J.-P.;
 RT "Organization of a Clostridium thermocellum gene cluster encoding the cellulosomal scaffolding protein CIPa and a protein possibly involved in attachment of the cellulosome to the cell surface";
 RL J. Bacteriol. 175:1891-1899(1993).
 [3]
 RN X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 184-321.
 RP MEDLINE=97238934; PubMed=9083107;
 RX Shimon L.J.W., Bayer E.A., Morag E., Yaron S., Shoham Y., Frolov F.;
 RT "A cohesin domain from Clostridium thermocellum: the crystal structure provides new insights into cellulosome assembly";
 RL Structure 5:381-390(1997).
 [4]
 RN X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 368-522.
 RP MEDLINE=97076134; PubMed=8918451;
 RX Torno J., Lamed R., Chirino A.J., Morag E., Bayer E.A., Shoham Y., Steltz T.A.;
 RT "Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose";
 RL EMBO J. 15:5739-5751(1996).
 [5]
 RN X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1213-1361.
 RP MEDLINE=98022914; PubMed=9402065;
 RX Tavares G.A., Beguin P., Alzari P.M.;
 RT "The crystal structure of a type I cohesin domain at 1.7-A resolution";
 RL J. Mol. Biol. 273:701-713(1997).
 [6]
 RL J. FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE CELLULOLYTIC ENZYMES.
 CC CELLULOSE BINDING LOCATION: CELL SURFACE.
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE.
 CC -1- DOMAIN: CONTAINS 9 COPIES OF A DOMAIN (COHESIN) OF ABOUT 150 RESIDUES THAT BINDS TO THE DOCKERIN DOMAIN BORN BY THE CATALYTIC COMPONENTS OF THE CELLULOSE.
 CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS WELL AS IN OTHER C.THERMOCCELLUM CELLULOSE ENZYMES. THIS DOMAIN MAY FUNCTION AS THE BINDING LIGAND FOR THE S1 COMPONENT.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC -----NOT_ANNOTATED_CDS.
DR EMBL; X67506; CAA47840.1; -.
DR PIR; S36859; S36859.
DR PDB; 1ANU; 23-JUL-97.
DR PDB; 1A0H; 08-JUL-98.
DR PDB; 1NBC; 26-SEP-97.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR002102; Cohesin.
DR InterPro; IPR002105; Dockerin_1.
DR Pfam; PF00942; CBD_3; 1.
DR Pfam; PF00963; Cohesin; 9.
DR Pfam; PF00404; Dockerin; 1; 2.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00448; CLOS_CELLOUSOME_REPEAT; 2.
KW Cellulose degradation; Cell wall; Glycoprotein; Repeat; Signal;
KW 3D-structure.
FT SIGNAL 1 28
FT CHAIN 29 1853 CELLOUSOMAL SCAFFOLDING PROTEIN A.
FT DOMAIN 364 522 CELULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 54 1694 9 X 150 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 29 182 1-1.
FT REPEAT 183 322 1-2.
FT REPEAT 560 704 1-3.
FT REPEAT 724 866 1-4.
FT REPEAT 889 1031 1-5.
FT REPEAT 1054 1196 1-6.
FT REPEAT 1219 1361 1-7.
FT REPEAT 1384 1526 1-8.
FT REPEAT 1548 1690 1-9.
FT DOMAIN 1791 1846 2 X 24 AA APPROXIMATE REPEATS.
FT REPEAT 1791 1814 2-1.
FT REPEAT 1824 1847 2-2.
FT COMPLECT 1615 1615 A -> AA (IN REF. 1).
SQ SEQUENCE 1853 AA; 196831 MW; 3ABDDC03ABRC5372 CRC64;
Query Match 7.5%; Score 413; DB 1; Length 1853;
Best Local Similarity 22.3%; Pred. No. 3.2e-15;
Matches 263; Conservative 128; Mismatches 363; Indels 424; Gaps 59;
OY 119 LTTVDAALISQGGTPEIEIVYIDLPGRCAALASNEIPATAAGL----- 166
DB 14 LTTTAAAMPVVSAAATVET-----GKTAAGSKVEIITLKGVPSKGMANCEFVL 67
OY 167 ---OTVETQYIDPIASILSNPKYS-----LRTVITIEPDSLPN--AVTNMSI- 209
DB 68 GYDPAVLEETEVR--GSIHKDPPSKSPDSATYPRKMIYLFADDSGSGTATITQDGYF 126
OY 210 -----QACATAVPYIEQIEVALTKLHAIPNVIY--MDAAHSGWLGMPNNAS-GYV 258
DB 127 ATIVATVKSAAAPITLLEVQ-AFADNDLVEISTFVAGVNLGSSVPTQPNVPSDGVV 185
OY 259 QEOVKVLAASGVNIDGFTVTANTYPLKEPFMTATQOQVGPESANF-IQAMPD--- 314
DB 186 VEIGKV-TGSGV-----TVEIP--VYFKGVPSKGIANDCEVFRYDPNVLE 228
OY 315 ---IDEADVAVDLYSRLVAAGFPSSIGMLIDTLRNGMGSPPTGPSTVDTVNTFVNOGK 371
DB 229 IIGIDPGDIYD-----PN-----PTYSFDTAITPDKRI 257
OY 372 IDLRQHRGIMCNGAGLGQPPQASPTDFPNAHLDAYWIKPQ---ESDGTSAAD--- 425
DB 258 IYF-----LFAEDSGTG---AYAITKDGVPFAKIRATVKSAPGYITFDEVGAFDNDLV 308
OY 426 -----PTTGKSDPKMCDPYT-----TSYVLTNALPNSITACQMPFA 463
DB 309 EOKVSEFIDGGVNVGNATPRTGATPTNTATPTKSATATPRPSVPTNTPTNTP----- 360
OY 464 QFDQLVANAARPAVVGSGLKVOYKKNDSAPGDNQIKPGLQLVNTGSSVDLSVTVRYWMT 523

```

```

DB 361 -----AN---TPVSGNLKVEEYNSNPSTNSINPORKVNTWGSSALDLSKTLIRYYT 411
OY 524 RDGGSSTIYVNCDMWAM-----GCCNIRASGSVNPATPRADYLOLSEFTGTLAA 574
DB 412 VDQKDDQTFW-CDHAALISNGSYNGITSNVAKGTVEKWSSTNNADTYLEISFTGTLPE 470
OY 575 GGSSTGIQNRVAKSWSNSTETNDYSYGNITTFQWTKYTVVNVGLWGTED----- 627
DB 471 GAHV-QIOGRFAKNDSWNTQSDNYSFKSASQFVEMDQVATLNVGLWKGKGGGSVPS 529
OY 628 ---VTPSV-----PTGL-----VTVGSSSVSIAMNAS-TDNV 658
DB 530 TQPVTPTPATTPPATTKPATTPATTPSPDDPAIKIKVDVNAKPPDVTNIPRFSGISK 589
OY 659 GVAHYN-VY-----RNGVAVGQPTVTSFTDGL-----AAGT-AY 691
DB 590 GIANDCEVYSYDPNVLEIIEIKPGLIYDNPDKSFDFAVYDPRKIYVLFADDSGTGAY 649
OY 692 TYT-----VAAVDA-----AGNTSAPS 708
DB 650 AITKDGVEATIVAKKSGAPNGLSVIKFEVVGAFANNDLVEORTQFFDGVNVGDTVPPT 709
OY 709 TPVD--CFPGPNQNGVTSVQDEYRVTQTEWMSAQQCLTINTATGAMTVSTANPSGGTG 766
DB 710 TTTTPTPTPTDSNAV-----RIKVD-----TVAKPDYVRIYVRESG--- 748
OY 767 GAPATPYSIYKG---CHWGCNTKKNVGMPIQISQISAVT-----SMST----- 808
DB 749 -----IPS--KGIANCDFYSYDPNVLEIIEI-EPGDIYDNPDKSFDTAVYDPRKIIV 800
OY 809 -----QVSSGADVAYD--IWTNSTPTTGPNNGEIMIMLSRG-----GVQPG 852
DB 801 FLFAEDSGTGAATATKDGVEATIVAKKSGAPNGLSVIKFEVVGAFANNDLVEORTQFFD 860
OY 853 SQTATGVYAGTATWVWVWQOQTSWKIISYVLPFGATSIISLDLKA1----- 898
DB 861 GGVNVGDTTPPAT-----PTTPVTPPTTDDLDVARIKVDIVNAKPPGDT 904
OY 899 -----PADAAAGSLNITSY-----LDVAGEFIEWGGQGLGNSPSVSV--- 939
DB 905 VRIPIVRFSGIPSKGIAN-CDFYSYDPNVLEIIEIEPGDIYDNP--DKSFDTAVYD 960
OY 940 -----TSGVACRATYVNSDMGSGFTATYTVNTGSRATNGWTV-----AMSG 984
DB 961 KRIIVFLFAEDSG---TGAVAITKD--GVPAIT-VAKVKSAGAPNGLSVIKFEVVGGA 1012
OY 985 GNOQTVNTWNTALTOGASVATATNLSNVNIQPGSTI 1022
DB 1013 NNDLVEQ--KIQFPDGVNVGDTT-----EPATPTT 1041
RESULT 12
CIPB_CLOTM
ID CIPB_CLOTM STANDARD; PRT; 772 AA.
AC 001866;
DT 01-OCT-1994 (Rel. 30, Created)
DI 01-OCT-1994 (Rel. 30, Last sequence update)
DR 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cellulosomal scaffolding protein B (Cellulosomal glycoprotein SI/Slu)
GN CIPB.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS;
RA MEDLINE=93146373; PubMed=1490597;
RA Poole D.M., Morag E., Lamed R., Bayer E.A., Hazlewood G.P.,
RA Gilbert H.J.;
RT Identification of the cellulose-binding domain of the cellulosome
subunit SI from Clostridium thermocellum YS."

```

```

RL EMBL Microbiol. Lett. 78:181-186(1992).
CC -1- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT
CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE
CC CELLULOTIC ENZYMES PROBABLY THROUGH THE BINDING OF THE NINE
CC REPEATED DOMAINS WITH THE DUPLICATED SEGMENT OF 24 AMINO ACIDS
CC PRESENT IN CATALYTIC SUBUNITS OF THE CELLULOSE.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE.
CC -1- SIMILARITY: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X68233; CAA48312.1; -.
DR HSP: 006851; INCB.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR002102; Cohesin.
DR InterPro: IPR002105; Dockerin_1.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00963; Cohesin; 3.
DR Pfam: PF00404; Dockerin_1; 2.
DR ProDom: PD001947; CBD_3; 1.
DR ProSite: PS00018; EF_HAND; UNKNOWN_1.
DR ProSite: PS00448; CLOS_CELLULOSE_RPT; 2.
DR PROSITE, PS00448; CLOS_CELLULOSE_RPT; 2.
DR Cellulose degradation; Cell wall; Glycoprotein; Repeat.
KW NON_TER
FT DOMAIN <1 607 3 x 147 AA APPROXIMATE REPEATS.
FT REPEAT <1 80 1 (PARTIAL).
FT REPEAT 94 240 2.
FT REPEAT 462 607 3.
FT DOMAIN 81 93 PRO/THR-RICH.
FT DOMAIN 241 272 PRO/THR-RICH.
FT DOMAIN 273 439 CELLULOSE-BINDING.
FT DOMAIN 440 461 PRO/THR-RICH.
FT DOMAIN 710 766 2 x 24 AA APPROXIMATE REPEATS.
FT REPEAT 710 766 2-1.
FT REPEAT 743 766 2-2.
SO SEQUENCE 772 AA; 82491 MW; BBF06DE5E094FE10 CRC64;

Query Match 7.2%; Score 401; DB 1; Length 772;
Best Local Similarity 24.8%; Pred. No. 4,7e-15;
Matches 151; Conservative 88; Mismatches 213; Indels 156; Gaps 24;

OY 426 PTTGKSDPMDPPTTYS-----YGVLTALNPSPIAGQWPPAOFDQVYANARPAVVGSG 480
DB 238 PTKGAPPTNATPTKSAITAPGHSVPTPTPTNP-----AN---TPVSGN 280
OY 481 LKVOYKNNDAPGDNQIKRDLQVNTGSSVDSTVTVRYWFTRDGSGSTLVYNCDDMAAM 540
DB 281 LKVEFYNSNSDPTNSTNPFKNTGSSAIDSKILTKYTYVDQKQOTFW-CDHAAI 339
OY 541 -----GCGNIRASFGSVNATPTADTYTQLSTFTGLAAGSGTGEIQNRVKNKSDMS 591
DB 340 IGSNGSYNGITSNVKGFVAKMSSNTNADTYLEISFTGLTEGCAHV-OIQGFPAKNDMS 398
OY 592 NFEETNDYSGTNTTTFODWTKVTVYVNGVLVWG-----TEDVTPSPVPTGLVWG 641
DB 399 NYTOSNDISFKRSQFEMDQVYALNGVLWKGKEPGSVVPTSTPTTPPTKPPATP 458
OY 642 VSGGSVLAMNASTDNGVAHYNVRNGVLVGOPT-VTSTDTGLA-AGTATYTYAAND 699
DB 459 ITPSDPDPAIKIKVDYVA-----KPGDYVNIIVRFSGIPSKCIANCDVYSY----- 506
OY 700 AAGNTSADSTPYDCTPGRNQNGYTSYDGEERYQTENMSAQCCLTINTA----- 750
DB 507 -----DENVLEIETIKPEGLIVDPNPKS-----FDRAVYDPRKMTV 543

```

```

OY 751 -----TGAWTVSTANSGGTGAPATPSPYIKGCHWNCCTTKNVMG-----IQISQ 797
DB 544 ELPEDSGTGAVAIT-----EDGVFAIVAKVY-----SGAPEGFSALISE 565
OY 798 IG-----SAVTSNSTQVSSG-----AYDAVDIWTSTPTTGP----- 833
DB 586 FGAFAUNDVEVEETDLINGVLVTKRPVIEGKVGSGYLLPDPSPATAPALVYKGFWEI 645
OY 834 NGTEIMWINSRGVQPFSSGATATGTVAGHTVWVMOGQTSWKIISVYLPATGISNL 893
DB 646 VGETLAVTDANG-----YEITGVPANMAGTYLTKISRATYLD-RVIANVYVGTDSVST 700
OY 894 DLKAIK-ADMAARGSLNTSDYLLDYEAQFELMOGGGSGNSFSVSYT-SGCVACRAFTY 950
DB 701 QAPIMMVGDIYVDNSIN-----LLDYAEVIRCPNATKGSANVVEELDIRNAINMGM 756
OY 951 VVNSDMS 958
DB 757 IVKHFQA 764

RESULT 13
GUNA_PAEIA STANDARD; PRT; 700 AA.
ID GUNA_PAEIA
AC P29719;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase A) (EG-A).
GN CELA.
OS Paenibacillus laetus (Bacillus laetus).
OC Bacteria; Firmicutes; Bacilli; Clostridium group;
OC Bacillus/Staphylococcus group; Paenibacillus.
OX NCBI_TaxID=1401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PL236;
RX MEDLINE=92276330; PubMed=1592807;
RA Hansen C.K., Diderichsen B., Joergensen P.L.;
RT "Cella from Bacillus laetus PL236 encodes a novel cellulose-binding
RT endo-beta-1,4-glucanase."
RL J. Bacteriol. 174:3522-3531(1992).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- PTM: A SHORT FORM (EGA-S) ARISES FROM POSTTRANSLATIONAL
CC PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-L.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY J (FAMILY 44 OF GLYCOSYL
CC HYDROLASES).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: M76588; AAA22303.1; -.
DR PIR: B41897; B41897.
DR PIR: S27499; S27499.
DR HSP: 006851; INCB.
DR InterPro: IPR001956; CBD_3.
DR Pfam: PF00942; CBD_3; 1.
DR ProDom: PD001947; CBD_3; 1.
KW Cellulose degradation; Hydrolyase; Glycosidase; Signal.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 700 ENDOGLUCANASE A.
FT DOMAIN 548 700 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT SITE 213 213 BY SIMILARITY.
SO SEQUENCE 700 AA; 76910 MW; 3D5C8CADA53BE0F CRC64;

```


DB 466 QSGDYSEKSNTEKTTKTTLYDQSKLIWETE 497

Search completed: August 29, 2002, 16:38:45
Job time: 965 sec

0 0 0

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:37:47 ; Search time 125.36 Seconds
(without alignments)
1439.324 Million cell updates/sec

Title: US-09-917-384-6
Perfect score: 1 MERTQOSGRNCRVGRGTRM.....GFGSYSGTNAAPTLSCTAS 1043
Sequence: BLOSUM62
Gapop 10.0 , Gapext 0.5

Scoring table:

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP virology:*
16: SP bacteriophage:*
17: SP archaea:

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1239.5	22.4	683	16	Q9PDW2
2	1051	19.0	579	2	Q86730
3	1015.5	18.3	454	2	Q9KH72
4	1000	18.1	596	2	Q60029
5	722.5	13.1	351	2	Q9S3V4
6	705	12.7	490	3	Q96V97
7	694	12.5	491	3	Q96V98
8	649	11.7	377	2	Q08468
9	647.5	11.7	381	2	Q9RJY3
10	633.5	11.4	371	2	Q54331
11	629.5	11.2	382	2	Q9KH71
12	618.5	11.2	460	3	Q02321
13	597.5	10.8	444	3	Q96V02
14	589.5	10.6	443	3	Q9C1R4
15	581.5	10.5	436	3	Q9P8N1
16					Q9P8N1 coriolus ve

17	569	10.3	465	3	Q96TP4	Q96TP4 pleurotus s
18	565.5	10.2	476	3	Q9C1S9	Q9C1S9 humicola in
19	554.5	10.0	384	2	Q9X602	Q9X602 streptomyc
20	554	10.0	471	3	Q9HEX7	Q9HEX7 trichoderma
21	544	9.8	457	3	Q93837	Q93837 acromonium
22	499	9.0	472	3	Q93860	Q93860 piromyces r
23	495.5	9.0	930	2	Q9REX5	Q9REX5 caldicellul
24	482.5	8.7	921	2	Q9L8L8	Q9L8L8 caldicellul
25	457	8.4	1751	2	Q9AOC4	Q9AOC4 caldicellul
26	452	8.2	458	3	Q9UW11	Q9UW11 piromyces r
27	450.5	8.1	1711	2	Q9X3P6	Q9X3P6 caldicellul
28	449	8.1	1426	3	P78721	P78721 orpionmyces
29	444.5	8.0	449	3	Q9X3P6	Q9X3P6 caldicellul
30	443	8.0	460	3	Q9UW10	Q9UW10 piromyces r
31	436	7.9	428	3	Q12646	Q12646 neocallitmas
32	434	7.8	170	2	Q9REX6	Q9REX6 caldicellul
33	434	7.8	1770	2	Q9X3P5	Q9X3P5 caldicellul
34	431.5	7.8	1779	2	Q52374	Q52374 caldicellul
35	421	7.6	1915	2	Q9P8L0	Q9P8L0 acetylviro
36	417	7.5	376	3	Q9P8L0	Q9P8L0 piromyces r
37	412	7.4	459	3	P78720	P78720 orpionmyces
38	410	7.4	1000	2	Q24820	Q24820 thermophil
39	407.5	7.4	260	2	Q33897	Q33897 rhodothermu
40	402	7.3	997	2	Q92411	Q92411 bacillus su
41	386.5	6.9	2316	2	Q9FDJ9	Q9FDJ9 bacteroides
42	384.5	6.9	890	2	Q86727	Q86727 streptomyc
43	383	6.9	499	2	Q45532	Q45532 bacillus su
44	381	6.9	508	2	Q93LD0	Q93LD0 bacillus su
45	380	6.9	499	2	Q93TD6	Q93TD6 bacillus su

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	683 AA.
Q9PDW2	Q9PDW2			
AC	Q9PDW2			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	1,4-BETA-CELLULOBIOSIDASE.			
GN	XF1267.			
OS	Xyloella fastidiosa.			
OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;			
OC	Xyloella.			
OX	NCBI_TaxID=2371:			
RP	[1]			
RC	SEQUENCE FROM N.A.			
RC	SPRAIN-9A5C:			
RC	MEDLINE-20365717; PubMed-10910347;			
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,			
RA	Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,			
RA	Bueno M.R.P., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,			
RA	Colautino N.B., Colombo A.A., Camargo L.E.A., Carraro D.M., Carer H.,			
RA	Costa L., Costa F.F., Costa M.C.R., Costa-Neto C.M.,			
RA	Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,			
RA	Faccinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,			
RA	Fraga J.S., Franca S.C., Franco M.C., Fromme M., Furian L.R.,			
RA	Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,			
RA	Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,			
RA	Krieger J.E., Lemos M.V.F., Laisret F., Lambais M.R., Leite L.C.C.,			
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,			
RA	Machado M.A., Madeira E.A.L., Martins E.M.F., Matsukuma A.Y.,			
RA	Marques C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,			
RA	Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,			
RA	Moore D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,			
RA	Nhal A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,			
RA	de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Paris A.,			
RA	Peixoto R.B., Pereira G.A.G., Pereira H.A., Jr., Pesquero J.B.,			
RA	Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,			
RA	de Rosa V.E., Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,			
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A., Jr.,			


```

Db      265 KTEYIDPIALIDSDSKYADRIYTVIEIDSLPLVTVNNGRPTAFENCVMKANGYQK 324
      223 IEVALTKLHAIPNVYIYMDAAHSGMLGMPNNAAGYVOEVOKYLNASIG--VAGIDGFTN 280
      325 VGVALNMLGAVGVNYYNVDAGHHGMLGMDSN-FGPSAEIEFKPAATEGATLDDVHGFVN 383
      281 TANTYPLKEPMTATQVGGQPVESANFYQWNPDIIDEADYADLYASRLYAAGFPSSIGML 340
      384 TANSALKKEENFKITDSDVNGTSVQSDVDMNOYTDLSIAQAMRKLVSLGFDQMLGML 443
      341 IDTLRNGGSGNEPTGPSTADVTNFEVQSKIDLROHRLGMCNONGAGLQGPQASPTDF 400
      444 IDTSNNGGAGDRPTGPGATDVNTVYNGGRYDRIRHLGMCNOSGAGLGERQASPA-- 501
      401 PNAHLDAYVWIKPGESEGTSAASDPTTGKSKSDPMCDPTT--TSYGVLTNALPNSPIA 457
      502 --AGIDAYVWMMKPGESEGSKSLIDNPDGKGFDRMCDPTTGTGNERNNGNSGALPDAPIS 559
      458 GMPFPAQFDOLVANARP 474
      560 GAMFSHQFOELMKNAYP 576

```

```

RESULT 3
ID      09KH72      PRELIMINARY;      PRT;      454 AA.
AC      09KH72;
DT      01-OCT-2000 (TREMBlrel. 15, Created)
DT      01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT      01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE      CELLULOSE PRECURSOR.
GN      EX.
OS      Thermomonospora fusca.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Streptosporangineae; Thermomonosporaceae;
OC      Thermobifida.
OX      NCBI_TaxID=2021;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Al Y.-C., Wilson D.B.;
RT      "Genomic DNA sequence encoding Thermomonospora fusca cellulase with
      overexpression."
RL      Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF288074; AAF75786.1; -.
DR      HSSP; P07987; ICB2.
DR      InterPro; IPR001524; Glyco_hydro_6.
DR      Pfam; PF01341; Glyco_hydro_6; 1.
DR      PRINTS; PR00733; GLHYDRLASE6.
DR      ProDom; PD003733; Glyco_hydro_6; 2.
DR      ProSite; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
KW      SIGNAL.
FT      CHAIN 1 31 POTENTIAL.
FT      SIGNAL 32 454 CELLULOSE.
SQ      SEQUENCE 454 AA; 48624 MW; F8CDIC4CD4807134 CRC64;

```

```

Query Match      18.3%; Score 1015.5; DB 2; Length 454;
Best Local Similarity 47.3%; Pred. No. 1.le-43;
Matches 221; Conservative 60; Mismatches 147; Indels 39; Gaps 12;

      27 LRAVLGAVSIASTIVPLAMQHPATIAH-----VDNPAGATFPVNVYMAOEVOSEAN 82
      6 LRA--LLGAAA--AAVSAALAFPPSQAAANPGEKVDNPEGAKLYVNVWSKAAAEFG 62
      83 QTNATLAAKRVYSTSTAVMMDRIAING---VNGSGPLTTYLDALISQOQGTPEYI 138
      63 S-----AVANESTAVWMDRIGALIGNSPTGSMGLRDHLEAV-RQSGDPLTI 111
      139 EIVYIDPGRDCAALASNGELPATAAGLQTYEYIDPASTLSN-PRYSRLRYITVIEP 197
      112 QVVIYVNLPGRDCAALASNGELGPDE--LDRYKSEYIDPADIWMDYENLRIVAILIEI 169

```

```

      198 DSLPNAVNMNSIOACATAVY-----YEOGIEYALTKLHAIPNVYIYMDAAHSGMLGMP 251
      170 DSLPLVTVNNGGNGTGLCAVYMKONGYGVYALARKGELIPNVYNIIDAAHGMIGMD 229
      252 NNAAGYVOEVOKYLNAS-IGVNGIDGFTVNTANTYPLKEPMTATQVGGQPVESANFYQ 310
      220 SNFPSVDIEFQAANMAGSTVDYVHGFTSNTANTSATVEYLDVNGTVNGQLIRGSKVD 289
      311 WNPDIIDEADYADLYSRLYAAGFPSSIGMLIDTLRNGGSGNEPTGPSTADVTNFEVQSK 370
      290 WNOYVDELSTFQDLRLQALINKGRFSIDIGMLIDTSRNGMGSGPNRPTGSSSTDLNTYVDS 349
      371 KIDLROHRLGMCNONGAGLQGPQASPTDFPNAHLDAYVWIKPGESEGTSAASDPTTGK 430
      350 RIDRRIHGNMNCNONGAGLGERPTVNPV---PGVDAYVWMMKPGESEGSSEELPNDGEK 405
      431 KSDPMCDPTTTSYGVLTN---ALPNSPIAQMPFPAQFDOLVANARP 474
      406 GFDRMCDPTTQGNANNGNPPSALPNAPISGHWFSAQFRELANARP 452

```

```

RESULT 4
ID      060029      PRELIMINARY;      PRT;      596 AA.
AC      060029;
DT      01-NOV-1996 (TREMBlrel. 01, Created)
DT      01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT      01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE      BETA-1,4-EXOCYLLULOSE PRECURSOR (EC 3.2.1.91).
GN      E3.
OS      Thermomonospora fusca.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Streptosporangineae; Thermomonosporaceae;
OC      Thermobifida.
OX      NCBI_TaxID=2021;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Zhang S., Lao G., Wilson D.B.;
RT      "Characterization of a Thermomonospora fusca exocellulase."
RL      Biochemistry 34:0-0(1995).
DR      EMBL; U18978; AAG62211.1; -.
DR      HSSP; P07986; IEXG.
DR      InterPro; IPR001919; CBD_2.
DR      InterPro; IPR001524; Glyco_hydro_6.
DR      Pfam; PF00553; CBD_2; 1.
DR      Pfam; PF01341; Glyco_hydro_6; 1.
DR      PRINTS; PR00733; GLHYDRLASE6.
DR      ProDom; PD003733; Glyco_hydro_6; 2.
DR      ProSite; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
KW      SIGNAL; Hydrolyase; Glycosidase.
FT      CHAIN 1 38 POTENTIAL.
FT      SIGNAL 39 596 BETA-1,4-EXOCYLLULOSE.
SQ      SEQUENCE 596 AA; 63547 MW; BOFA5277FE7721E0 CRC64;

```

```

Query Match      18.1%; Score 1000; DB 2; Length 596;
Best Local Similarity 47.7%; Pred. No. 9.3e-43;
Matches 210; Conservative 55; Mismatches 143; Indels 32; Gaps 9;

      50 PALAATHVNDPYGATFPVNPYMAOEVOSEANQTNATLAAKRVYSTSTAVMMDRIAA 109
      172 PTNNGEVNDNPFSGAKLYVNPWSAKAAEPGS-----AVANESTAVWMDRIGA 221
      110 ING---VNGSGPLTTYLDALISQOQGTPEYIYIDLPGRDCAALASNGELPATAAG 165
      222 IEGNDSPTGSMGLRDHLEAV-RQSGDPLTIQVVIYVNLPGRDCAALASNGELGPDE-- 278
      166 LQTEYIDPISILSN-PRYSRLRYITIEPDSLPAVNMNSIOACATAVY----- 218
      279 LDRYKSEYIDPADIWMDYENLRIVAILIEIDSLPLVTVNNGGNGTGLCAVYMKONG 338
      219 YEOGIEYALTKLHAIPNVYIYMDAAHSGMLGMPNNAAGYVOEVOKYLNAS-IGVNGIDG 277

```

```
DB 339 YNCGVGLARKGEIENPVNYIDAHHGMISNCGSPDIPIEAMNSGSIYDVYHGF 338
QY 278 VYNTANTYPLKEPMTATQOVGQPVESANFYQWNPDIIDEADYAVDLYSRLVAAGPSSI 337
DB 399 ISNTANYSATVEBYLDVNGTNGQLIRQSKWMDWNOYDELSPVODLRQALIKGFRSDI 458
QY 338 GMLIDIRLNGMGSPNPTGSPSTADVTNTFYVNOCKIDLROHRLGIMCQONAGLGCPQASP 337
DB 459 GMLIDISRNMGWGPNNRPTGSSSTIDNTYVESRIDRRIHPGMCNQAAGLGEPTVNP 518
QY 398 TPEPNAHLDAYWIKPBGESDGTSAASDPPTGKSDPMCDPTVYTSYGVLTN---ALPNS 454
DB 519 A----PGVAYVWVKPRGSDGSEELPNDEGKGFPRMCDPTVQGNARNONNSGALPNA 574
QY 455 PIRAGWFPAGFQDOLVANARP 474
DB 575 PISGHWFSAQFRELANAVP 594

RESULT 5
Q9S3V4 PRELIMINARY; PRT; 351 AA.
Q9S3V4
AC Q9S3V4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 1,4-BETA-CELLULOBIOLYDOLASE A (EC 3.2.1.91) (FRAGMENT).
GN CFLA.
OS Cellulomonas flavigena.
OC Bacteria; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1711;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDB8531;
RA Herrera A., Gutierrez A., Salgado L.M., Ponce-Neoyola T.;
RT "Molecular characterization of cellulases from Cellulomonas
RT flavigena.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBD databases.
DR EMBL; AF17344; AAD48493.1; -.
DR HSSP; P07987; 1CB2.
DR InterPro; IPR001524; Glyco_hydro_6.
DR Pfam; PF01341; Glyco_hydro_6; 1.
DR PRINTS; PR00733; GLHTRKLA56.
DR PRODOM; PD00373; Glyco_hydro_6; 1.
DR PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
DR KMW HYDROLASE; Glycosidase.
FT NON_TER 351 351
SO SEQUENCE 351 AA; 37615 MW; BCDE63B5DF335C2 CRC64;

Query Match 13.1%; Score 722.5; DB 2; Length 351;
Best Local Similarity 48.1%; Pred. No. 4.4e-29;
Matches 166; Conservative 45; Mismatches 105; Indels 29; Gaps 13;

QY 21 PASKRLRAGVLAGANVSIASIT-VPLAMQ--HPAIAA-THVDNPIYGATFTFVNPYNA-OE 75
DB 6 PSYARRRRRTV--GAFTTALAVALVPLTSLFSLSPASAAEAEKRVDPYAGAVOYVNPNTAASS 63
QY 76 VOSEANQTNATLAARKRVYSTYSTAVMMDRIAINGVNGPGLTLYTLDAALSOOGGT-T 134
DB 64 VARSAKQSDASIAAKMQVYAKOPTAVMMDRISAITGNADGKGLKTHLDAAVQOQKAAK 133
QY 135 PEYIEIYVDLPGRDCAALASNGELPATAGLQTYETQYIDPITASILSNPKYSLRIVTI 194
DB 124 PLVFNLVYINLPGRCDCYALASNGELPATDAGLARQSEYIDPIVALLGKREYSNLRIAAT 183
QY 195 IEPDSLPNAVNTNNSIOACATAVPYIEQGLEFALTKLHALPNVITYYMDAAHSGMLGPNNA 254
DB 184 IEPDSLPLNLTNISASCOQAAPIYRQGVKYLADKLALGNVNYIDIGISGMLGMDSNA 243
QY 255 SGVYQ---EVQKVLNLSIGVNGIDGFVTNTANTPLKE--PEWTAHQOYVGGQPVESAN-F 308
```

```
DB 244 GPAKIFAEYAKTTNA--GPASTITGFVSDVANTTPLEEPYVPERPSREGNPIREQYV 301
QY 309 YQWNPDIIDEAD--YAVDLY--SRLVAAGPSSIGMLIDTLRNGMG 349
DB 302 TSGNNELSTNRNLRYGATLHRVGSFGRGPE-----RRPWG 337

RESULT 6
Q96V97 PRELIMINARY; PRT; 490 AA.
Q96V97
AC Q96V97;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELLULOBIOLYDOLASE II-LIKE CELLULASE CELI.
GN CELI.
OS Orphiomyces sp. PC-2.
OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
OC Neocallimastacaceae; Orphiomyces.
OX NCBI_TaxID=50059;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PC-2;
RX MEDLINE=21405733; PubMed=11514516;
RA Steendackers P.J.M., Li X.L., Ximenes E.A., Arts J.G., Chen H.,
RA Ljungdahl L.G., Op Den Camp H.J.M.;
RT "Noncatalytic docking domains of cellulosomes of anaerobic fungi.";
RL J. Bacteriol. 183:5325-5333(2001).
DR EMBL; AF177205; AAL01212.1; -.
KW Hydrolyase.
SO SEQUENCE 490 AA; 54051 MW; 85F40D1280759886 CRC64;

Query Match 12.7%; Score 705; DB 3; Length 490;
Best Local Similarity 39.0%; Pred. No. 5e-28;
Matches 167; Conservative 54; Mismatches 147; Indels 60; Gaps 12;

QY 50 PAIAATHVDNPNVYAGATFFVNPYNAOEVOSEANQTNATLAARKRVYSTYSTAVMMDRIA 109
DB 118 PGEQYTHSGNPFSGVEFFLNPIYVAEYDAIEQMTNSLSLAKAKKKKTYSNALIMDTIK- 176
QY 110 INGVNGPGLTLYTLDAALSOOGCTTEBYEIT-VIYDLPGRDCAALASNGELPATAGLQ 168
DB 177 ----NMHQLWLETKGALQOQEFSGSKVLVFEVYDLPGRDCHALASNGELANDSDAQR 232
QY 169 YETQYIDPILASISNPKYSLRIVTITIEPDSLNAVYTNM-SIOACATAVPYIEQGLEVAL 227
DB 233 YKTEYIDVIEEKLK--YKQAPVYLLIVERPSLNLVTLNLDSTPACRDSKTYTLDGHAAYLI 290
QY 228 TKLHAIPNVYIYMDAAHSGMLGMPNNAAGVYQVOKVLNASIGVNG-IDGFVTNTANTYP 286
DB 291 OKLGVLPHVAMLYDIGHAFWLGMDNREKAGKYKVI--SSGPKGVNRGFTDNVANVTP 348
QY 287 LKPEPMTATQOVGQPVESANFYQWNPDIIDEADYAVDLYSRLVAAGPSSIGMLIDTLRN 346
DB 349 WEDEPTLSRGPEI-----EWNPCDPEKRYTLELMKDKRAGIESVYVEVCDTSRN 397
QY 347 GMGCPNEPTGSPSTADVTNPFVNOCKIDLROHRLGIMCQONAGLGCPQASPFTDPNNAHL 406
DB 398 G-----HKYD-RKHPEKMCNQTGVSIGARQASPVSGMD-YLD 433
QY 407 AYWIKPBGESDGTSAASDPPTGKSDPMCDPTVYTSYGVLTNALPNSPIAGWFPAGQPD 466
DB 434 AFYWIKPLGSDGTS---DTSARVDGYCGH-----DTAMKPAPRAGQWFOKHFE 480
QY 467 QIVANARP 474
DB 481 QGLENNAP 488

RESULT 7
Q96V96
```

ID	Q96V98	PRELIMINARY;	PRT;	491 AA.
AC	Q96V98			
DT	01-DEC-2001 (TREMBL,rel. 19, Created)			
DT	01-DEC-2001 (TREMBL,rel. 19, Last sequence update)			
DT	01-DEC-2001 (TREMBL,rel. 19, Last annotation update)			
DE	CELLULOSEHYDROLASE II-LIKE CELLULOSE CELH.			
GN	CELH.			
OS	Orpinomyces sp. PC-2.			
OC	Eukaryota; Fungi; Chytridiomycota; Neocallimastcales;			
OC	Neocallimastaceae; Orpinomyces.			
OX	NCBI_TaxID:50059;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-PC-2;			
RX	MEDLINE=21405133; PubMed=11514516;			
RA	Steenbakkers P.J.M., Li X.L., Ximenes E.A., Arts J.G., Chen H.,			
RA	Ljungdahl L.G., Op Den Camp H.J.M.;			
RT	'Noncatalytic docking domains of cellulosomes of anaerobic fungi.';			
RL	J. Bacteriol. 183:5325-5333(2001).			
DR	EMBL: AF177204; AAL01211.1; "			
KW	Hydrolase			
SQ	SEQUENCE 491 AA; 53956 MW; 7C7ACCC9BFF946C4 CRC64;			

Query Match	12.5%;	Score 694;	DB 3;	Length 491;
Best Local Similarity	38.6%;	Pred. No. 1.8e-27;		
Matches 165; Conservative	53;	Mismatches 149;	Indels 60;	Gaps 12

```

0Y PALAACHVNDPVGAFEEFVNPYMAOEYOSLAAOTNATLAAKRRVSTSTAVMMRIA 109
109 PPS : : : : : : : : : : : : : : : : : : : : : : : : : : : :
121 PGSOLTLSTGNPSEVEFFLNRPYYAAEDDAIAQMSNSSLAKAEKMTYENALIMLDTIK - 179
179 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 110 INGVNGSGPGLTYYDAALSOOGCTTPEVIEIVLYDLRPGRCDAALASNGELPATAGIQTY 169
169 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 180 ----NQOMLETNLKGLAAGTSGKVLGYFVVYDLRPGRCHALASNGELLANDSDQRY 235
235 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 170 ETOYIDPRLASTLSPKTSLSLRYITVITIEPDSLPRNAVYTM- SIOACATVAPYEOGIEALT 228
228 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 236 KTEYIDVIEEKIK--YKKSOPVLLIEBDSLANTVLENTPRACRDBEQIYLDOSHAYLIK 293
293 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 229 KLHAIIPNYIYMDAASHGLMPNNASGYOENOKVLNASTGVNG-IDGEVYNTANTYPL 287
287 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 294 KRGVLEPHAMLIDIGHAIVLMDORREKAARVYKSVI--SSGSPGKVGIEDVNAVANTPW 351
351 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 288 KEPEMTATQOYGGQRPVESANFYQNNPDIIDEADYANDLXSRUVAGPSSIGMLIDITLIRNG 347
347 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 352 EDRPLSRPBDT-----EMNRPCEDEKRYLEAMKDKRAKAI--SSVYFVSDTISNG 399
399 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 348 WGRPRNRPNGSPATDVNTFVNOSKIDLQHRGLMCONGAGIGORPOASPTDFENALIDA 407
407 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 400 -----HKRD-RKHPRCENONQGVGICARPOANPISMSD-YLDA 435
435 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 408 YVMIKRPGESDGTSAASDPTTGKSKSDPMKDPYITTSYGLVLTNALNSPACQMPAOFDO 467
467 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 436 FYWIKRLGSDGTS-----DTSARYDGYCGH-----ETAMKPARAAGQMPQKHFRQ 482
482 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 468 LVANARP 474
474 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 483 GLENANP 489
489 : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT	8			
008468				
ID	008468	PRELIMINARY;	PRT;	377 AA.
AC	008468;			
DT	01-JUL-1997 (TrEMBLrel. 04, Created)			
DT	01-JUL-1997 (TrEMBLrel. 04, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	CEL2 (EC 3.2.1.4).			
GN	CEL2.			
OS	Streptomyces halstedii.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;			
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.			

ON NCBI_TaxID=1944;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Jm8.
 RX MEDLINE=97307849; PubMed=9182697;
 RA Gada-Salas A.L., Fernandez-Abalos J.M., Sanchez P., Ruiz-Arribas A.,
 RA Santamaría-Sánchez R.I.;
 RT "Two genes encoding an endoglucanase and a cellulose-binding protein
 RT are clustered and co-regulated by a TTA codon in Streptomyces
 RT halstedii Jm8.";
 RL Biochem. J. 324:403-411(1997).
 DR EMBL; U51222; AAC45429.1; -.
 DR HSSP; P07986; 1EXG.
 DR InterPro; IPR001919; CBD_2.
 DR InterPro; IPR002594; Glyco_hydro_12.
 DR Pfam; PF00553; CBD_2; 1.
 DR Pfam; PF01670; Glyco_hydro_12; 1.
 DR Pfam; PF0004316; Glyco_hydro_12; 1.
 DR Pfam; PF0004316; Glyco_hydro_12; 1.
 KW Hydrolyse: glycosidase.
 SQ SEQUENCE 377 AA; 38829 MW; 9892191B741EC280 CRC64;

Query Match	11.7%;	Score 649;	DB 2;	Length 377;
Best Local Similarity	41.7%;	Pred. No. 2.4e-25;		
Matches 149;	Conservative 45;	Mismatches 137;	Indels 26;	Gaps 10

[illegible]

RESULT	9	
09RJY3		
ID	09RJY3	PRELIMINARY; PRT; 381 AA.
AC	09RJY3	
DT	01-MAY-2000 (TREMBLrel. 13, Created)	
DT	01-MAY-2000 (TREMBLrel. 13, last sequence update)	
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)	
DE	POTATIVE SECRETED CELLULASE B.	
GN	CE1B.	
OS	Streptomyces coelicolor.	
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.	
OX	NCBI_taxid:1902;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-A3(2);	
RA	Oliver K., Harris D.;	
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-A3(2);	
RA	Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;	

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE-97000351; PubMed-8843436;
 RA Redenbach M., Kleser H.M., Denaplatte D., Eichner A., Cullum J.,
 Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL; AL133210; CAB61599.1; -
 DR HSSP; P07986; 1EXG.
 DR InterPro: IPR001919; CBD_2.
 DR InterPro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR Prodom: PD004316; Glyco_hydro_12; 1.
 DR PROSITE: PS00561; CBD_BACTERIAL; UNKNOWN_1.
 SQ SEQUENCE 381 AA; 39199 MW; 97CD8FF56679E4EC CRC64;

Query Match 11.7%; Score 647.5; DB 2; Length 381;
 Best Local Similarity 40.7%; Pred. No. 2.9e-25;
 Matches 155; Conservative 45; Mismatches 144; Indels 37; Gaps 14;
 QY 684 GLAAG-----TAYTYVAADVDAAGNTSAPSTPYDCTPGPNONGVTSVODGEXRVQTNEMN 738
 DB 13 GLLAALGAVLAFAVLSLVTAAPAOADT--ICEP----FGTTTIQ-GRYVQNNRWG 65
 QY 739 SSAOOCITINTGAMTVSTANFSGGTGAPATPSIYKCHMGKCTTKNVGMPIDISOI 798
 DB 66 STATQCVTA-TDTG-FRVTQADGSAPTNGAPKPSYFVFGCHYTN-SGTCLPVRLLDV 122
 QY 799 GSAVTSWSTQVSSGAYDAVDIWTNSTPTTGQPGNGTEIMIMLNSRGVQPGFSQTATG 858
 DB 123 SAAPSISYSGFDGAYVNASYDIWIDLPARTDG-VNQTETIMTFNNGVPIQIGSPVGT- 180
 QY 859 VIVAGHTMNVWOGQTSMTIISVLTTPGATSIISNDLKAIFDAAARGSLNTSDYLLDVE 918
 DB 181 ASVGGRTMFWWSGGNSNDVLSFV-APSAISGMSFVDMDFVATVARGLAENDWYLTSTVQ 239
 QY 919 AGFEIMOGGOGIGSNSFSYSV---TSGGV-----ACRATYVNSDMGSGFTATVTV 966
 DB 240 AGFEPMQNGAGLAVNSFSSTVEGTGPGTDPGDPSGACAVSYGTNV-WQDGFADVTV 298
 QY 967 TWTGSRATGWTAVNSFGNGQVTNTWNTALTQSGAVTATNLYSNNVTQPGQSTTGFEN 1026
 DB 299 TWTGAPVDGQWOLAFPLPGQRITNANMASLTPSSGCVTATGASHNARIAPGSLSFQ 358
 QY 1027 GSYSGTNAAPT-----LSCT 1041
 DB 359 GTYGGAFAEPTGFRLLNGTACT 379

RESULT 10
 ID 054331 PRELIMINARY: PRT: 381 AA.
 AC 054331.
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CELLULASE B PRECURSOR.
 GN CELB.
 OS Streptomyces lividans.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID-1916;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-66;
 RX MEDLINE-94288649; PubMed-8017952;
 RA Witmann S., Shareck F., Kluempel D., Morosoli R.;
 RT "Purification and characterization of the CelB endoglucanase from

RT Streptomyces lividans 66 and DNA sequence of the encoding gene.";
 RL Appl. Environ. Microbiol. 60:1701-1703(1994).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-66;
 RA Shareck F.;
 RT Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U04629; AAB1950.1; -
 DR HSSP; P07986; 1EXG.
 DR InterPro: IPR001919; CBD_2.
 DR InterPro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF00553; CBD_2; 1.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR Prodom: PD004316; Glyco_hydro_12; 1.
 DR PROSITE: PS00561; CBD_BACTERIAL; UNKNOWN_1.
 KW SIGNAL.
 FT CHAIN 1 40 POTENTIAL.
 FT CHAIN 41 381 CELLULASE B.
 SQ SEQUENCE 381 AA; 39239 MW; A7E99BF590FA24EC CRC64;

Query Match 11.7%; Score 646.5; DB 2; Length 381;
 Best Local Similarity 40.7%; Pred. No. 3.2e-25;
 Matches 155; Conservative 45; Mismatches 144; Indels 37; Gaps 14;
 QY 684 GLAAG-----TAYTYVAADVDAAGNTSAPSTPYDCTPGPNONGVTSVODGEXRVQTNEMN 738
 DB 13 GLLAALGAVLAFAVLSLVTAAPAOADT--ICEP----FGTTTIQ-GRYVQNNRWG 65
 QY 739 SSAOOCITINTGAMTVSTANFSGGTGAPATPSIYKCHMGKCTTKNVGMPIDISOI 798
 DB 66 STATQCVTA-TDTG-FRVTQADGSAPTNGAPKPSYFVFGCHYTN-SGTCLPVRLLDV 122
 QY 799 GSAVTSWSTQVSSGAYDAVDIWTNSTPTTGQPGNGTEIMIMLNSRGVQPGFSQTATG 858
 DB 123 SAAPSISYSGFDGAYVNASYDIWIDLPARTDG-VNQTETIMTFNNGVPIQIGSPVGT- 180
 QY 859 VIVAGHTMNVWOGQTSMTIISVLTTPGATSIISNDLKAIFDAAARGSLNTSDYLLDVE 918
 DB 181 ASVGGRTMFWWSGGNSNDVLSFV-APSAISGMSFVDMDFVATVARGLAENDWYLTSTVQ 239
 QY 919 AGFEIMOGGOGIGSNSFSYSV---TSGGV-----ACRATYVNSDMGSGFTATVTV 966
 DB 240 AGFEPMQNGAGLAVNSFSSTVEGTGPGTDPGDPSGACAVSYGTNV-WQDGFADVTV 298
 QY 967 TWTGSRATGWTAVNSFGNGQVTNTWNTALTQSGAVTATNLYSNNVTQPGQSTTGFEN 1026
 DB 299 TWTGAPVDGQWOLAFPLPGQRITNANMASLTPSSGCVTATGASHNARIAPGSLSFQ 358
 QY 1027 GSYSGTNAAPT-----LSCT 1041
 DB 359 GTYGGAFAEPTGFRLLNGTACT 379

RESULT 11
 ID 09KITH1 PRELIMINARY: PRT: 371 AA.
 AC 09KITH1.
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CELLULASE 12A.
 GN CEL12A.
 OS Streptomyces sp. 11A68.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID-133452;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-11A68;
 RA van Solingen P., Meijer D., van der Kleij W.A.H., Barnett C.C.,
 Bolle R., Power S.D., Jones B.E.;
 RT "Cloning and expression of an endocellulase gene from a novel

Query Match	Similarity	11.2%	Score 618.5	DB 3	Length 460
Best Local Similarity	36.8%	Pred. No. 1e-23			
Matches 154	Conservative 55	Mismatches 144	Indels 65	Gaps 12	
QY	58	DNPYGATFEFVNPYVAOGESEANQTNATLAAKRRVYSTYSTAVWMRIAINGVNGP	117		
Db	105	NNPWTGFGQIFLSPRYANIEVAAAKQITDPTLSKKAASVANITPTFWLDSVAKI-----P	158		
QY	118	GLTYFLDAALSOOGCT-TPEVIEIYIYDIPGRDCAALASNGELPRTAAGLOTETQYIDP	176		
Db	159	DGTYLASASAGKSTGTQKOLVOIYIYDLPDDCDCAKANSGEFSIANNQANVE--YIDQ	217		
QY	177	IASISNPKYSLLVTYITPDLSLNAYNMSTIOCATVPRYEGGIEVALTKLAIPNV	236		
Db	218	IYAQIQ--QFPYRVAYAYIEPDSLNLNTLNQVCANAKTTYILACVNTLNLAKV--GV	274		
QY	237	YIYMDAHSGLWGPNNASGVYQVQKVLNASTIGVNGIDGFEYTNANTYPLKEPFTATQ	296		
Db	275	YMYMDAGHAGMLGPNLSPANQALPTQVWQNGKSPFIKGLATNVAN-----NALQ	326		
QY	297	QVGGQPVESANTYQWNPIDDEADYAVDLXSLVAAGFSSITMLIDTLRNGSGPNEPTG	356		
Db	327	RAASPPIT-----QGNPNDEIHYINALAPLLQOAGWAT--FIVDGRSG-----	370		
QY	357	PSTADVNTFVYQSKIDLEHQRGLCNONGAGLGGPQASPTDFPAHLIDAVYWKIPGE	416		
Db	371	-----VQNIHQGMDCMNKIGAGFGIRP---TNTGSGQIFDSTIYVWYKPGGE	413		
QY	417	SDGTSASDPPTGKKSDDPCDPTYYTTSYGVLTNALPNSPIAGQWPPAOFDOLVANNRP	474		
Db	414	CDGTSSNSSSP-----RYDSTCS-----LPDAQAPAEAGTWPQAYPQTLVSAANP	458		
RESULT	14				
ID	096VU2	PRELIMINARY;	PRT;	444	AA.
AC	096VU2;				
DT	01-DEC-2001 (Tremblrel. 19, Created)				
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)				
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)				
DE	CELLULOSE CEL6B.				
GN	CEL6B.				
OS	Lentilina edodes (Shiitake mushroom) (Lentinus edodes).				
OC	Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;				
OC	Agaricales; Tricholomataceae; Lentinula.				
OX	NCBI_TaxID=5353;				
RR	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=STAMETS CS-2;				
RA	Lee C.C., Wong D.W.S., Robertson G.H.;				
RT	"Cloning and characterization of two cellulase genes from Lentinula edodes.";				
RL	Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.				
SR	EMBL; AF411251; AAK95564.1; -				
SO	SEQUENCE 444 AA; 46368 MW; A08BEDD0835FE47C5 CRC64;				

[illegible][illegible]

Query Match	10.6%	Score 589.5;	DB: 3;	Length 443;
Best Local Similarity	33.4%;	Pred. No. 2.9e-22;		
Matches 158;	Conservative 65;	Mismatches 173;	Indels 77;	Gaps 16;
OY	7	SGRNRGRTGTTMPAISKRLRAGLAGAVSINASIVPLAMQHPAIATHTYDNYPAGATF	66	
Db	41	SCATC-----TVNAVYYSQCLPGSASAPRPISTSIGCGITTTSSAPSGTGTTTPAAGNPF	94	
OY	67	-----FVNPMQVQSEANQTNATLAKMRVSTYSTAVWMDRIALINGVGGPGLTYY	122	
Db	95	TEQIYLSPYANIEIAAATQISDPPTAAAKAAKVAIPFEIWDQVAKV-----PDLGTY	148	
OY	123	L-DALISQOQGTTPVIEIETVDLGGRCALANSGELPRAPAGQGTETQYIPIASIL	181	
Db	149	LADASKQSESKNYLVQIVYVDLPDRCCALANSGEFTIADNGEANYH-DTIDIVAQI	207	
OY	182	SNPKYSSRLIVYIIIPEDSLPNVATNMSIQACATAPVYEGQIEYALTLHAIPNVYIYND	241	

```

Db      208 K--QYDPVHVAVIIEPPDSLANTLVNLSVAKCANAQTTTECVYAMQQLSAV--GVTMYILD 264
QY      242 AAHSGWLGMPNNASGYVOEYOKVLNASIGVNGIDGFTNTANTYPLKEPFMTATQOYGQ 301
Db      265 AGHAGWLGMPNNASGYVOEYOKVLNASIGVNGIDGFTNTANTYPLKEPFMTATQOYGQ 316
QY      302 PVESANFYQWNPDIADYAVDLYSRLVAGFPSSIGMLIDTLRNGWGGPNEPTGPSTAT 361
Db      317 PTF-----QGDPMYDEMLYIEALAPLL--GSFPAH--FIVDQGRSG----- 353
QY      362 DVNTEYNQSKIDKROHRLGMLCNGAGLGQPPQASPTDFPNAHLDAYVWIKPPGESDGT 421
Db      354 -----VQDIRQWGMCMCNVLGAGFGTQP--TTNGSSLIDSIWVWPGGECDDGT 401
QY      422 AASDPTTGKSDPMCDPTTYSYGVLTNALPNSPIAGOMFPPOEDOLVANARP 474
Db      402 NTSSP-----RYDAHCG-----LPDATPNAPEAGTWFQATFETLVEKANP 441

```

Search completed: August 29, 2002, 16:37:51
 Job time: 961 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:22:35 ; Search time 47.58 Seconds
(without alignments)
535.433 Million cell updates/sec

Title: US-09-917-384-6
Perfect score: 5536
Sequence: 1 MERTQSGRNCRYGRGTTM.....GFGNGSYSGTNAAPTLCSTAS 1043

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5a_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5b_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5a_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/5b_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/5a_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/5b_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	646.5	11.7	381	4	US-09-216-295-22 Sequence 22, App1
2	639.5	11.6	352	4	US-09-286-691-26 Sequence 26, App1
3	639.5	11.6	352	4	US-09-687-147-26 Sequence 26, App1
4	637	11.5	386	4	US-09-321-981-5 Sequence 5, App1
5	633.5	11.4	371	4	US-09-104-308-1 Sequence 1, App1
6	633.5	11.4	371	4	US-09-321-981-1 Sequence 1, App1
7	564.5	10.2	312	4	US-09-216-295-21 Sequence 9, App1
8	556.5	10.1	461	3	US-09-118-319-9 Sequence 25, App1
9	555	10.0	360	4	US-09-286-691-25 Sequence 25, App1
10	555	10.0	360	4	US-09-687-147-25 Sequence 8, App1
11	548.5	9.9	457	3	US-09-118-319-8 Sequence 1, App1
12	544	9.8	470	3	US-09-142-759-1 Sequence 24, App1
13	538	9.7	360	4	US-09-286-691-24 Sequence 24, App1
14	538	9.7	360	4	US-09-687-147-24 Sequence 12, App1
15	538	9.7	360	4	US-09-687-147-24 Sequence 12, App1
16	538	9.7	360	4	US-09-687-147-24 Sequence 12, App1
17	538	9.7	360	4	US-09-687-147-24 Sequence 12, App1
18	467	8.4	1751	4	US-08-448-873-12 Sequence 12, App1
19	449	8.1	1426	4	US-08-448-873-12 Sequence 12, App1
20	444.5	8.0	449	4	US-09-136-574A-43 Sequence 43, App1
21	444.5	8.0	449	4	US-09-136-574A-43 Sequence 43, App1
22	444.5	8.0	449	4	US-09-136-574A-43 Sequence 43, App1
23	436	7.9	326	4	US-09-118-319-7 Sequence 7, App1
24	434.5	7.8	326	4	US-09-118-319-7 Sequence 7, App1
25	434.5	7.8	326	4	US-09-118-319-7 Sequence 7, App1
26	431.5	7.8	432	3	US-09-687-147-23 Sequence 23, App1
27	417.5	7.5	551	2	US-09-118-319-2 Sequence 2, App1

28	412	7.4	459	3	US-09-118-319-6 Sequence 6, App1
29	412	7.4	459	4	US-09-286-691-2 Sequence 2, App1
30	412	7.4	459	4	US-09-687-147-2 Sequence 2, App1
31	407.5	7.4	260	4	US-09-216-295-23 Sequence 23, App1
32	406	7.3	521	1	US-08-276-213-3 Sequence 3, App1
33	388.5	7.0	700	2	US-07-862-588B-2 Sequence 2, App1
34	376	6.8	616	4	US-09-136-574A-47 Sequence 47, App1
35	371.5	6.7	493	4	US-09-136-574A-47 Sequence 47, App1
36	371.5	6.7	493	4	US-09-136-574A-47 Sequence 47, App1
37	362	6.5	167	5	PCR-US95-13813-9 Sequence 9, App1
38	269.5	4.9	348	4	US-09-216-295-16 Sequence 16, App1
39	260.5	4.7	259	4	US-09-216-295-5 Sequence 5, App1
40	244.5	4.4	108	2	US-08-585-585A-2 Sequence 2, App1
41	244.5	4.4	108	2	US-08-249-037C-2 Sequence 2, App1
42	244.5	4.4	108	2	US-08-788-622B-2 Sequence 2, App1
43	244.5	4.4	108	3	US-08-788-621B-2 Sequence 2, App1
44	242.5	4.4	291	4	US-09-286-691-28 Sequence 28, App1
45	242.5	4.4	291	4	US-09-687-147-28 Sequence 28, App1

ALIGNMENTS

RESULT 1
US-09-216-295-22
; Sequence 22, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; TITLE OF INVENTION: No. 6268328el Variant EgIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Streptomyces lividans Celb
US-09-216-295-22

Query Match 11.7%; Score 646.5; DB 4; Length 381;
Best Local Similarity 40.7%; Pred. No. 3.9e-38;
Matches 155; Conservative 45; Mismatches 144; Indels 37; Gaps 14;

QY	684	GLAAG-----TAYTYVAAYDAAGNTSAPSTPYDCTPGFNONGVSYVDGGRVOTNEMN	738
DB	13	GLAALAGVALAFAVALVSLVTAAPAOADT--ICEP---FETTYIO-GRIVONNRMG	65
QY	739	SSAOCCLTIMATGAWTVSTANFSGGTGAPATPSIYKCHMGNCITRNVMQPIQISOI	798
DB	66	STAPQCVTA-TDVG-FEVTYQADSGAPTNKAPKSPYFENCHTNC-SPGTDLPVALDIV	122
QY	799	GSAYTSMSTQVSSGADVAVIWTNSTPTTQOPNGTEIMWLNSRGVQPPGSGTATG	858
DB	123	SAAPSSISYGVDDAVYANASYDIWLDPTARTDG-VNQTETIMFNRRVGPLOPGSPVGT-	180
QY	859	VTVAAGHTNVMQSGQOTSMKIISYLVLTGATGISLMDKALFADAAAGSINTSDYLDVE	918
DB	181	ASVGRTVEWVGSGNSGSDVLSFV-APSAISGMSFVDMDEVYATVANGLAENMWLYLSVQ	239
QY	919	AGFEIMOGGGLGNSFESVSV---TSGGV-----ACRATVYNSDMGSGFATVTV	966
DB	240	AGFEIMOGGAGLAVNSFESVETGTPGDTDPGDCPSACAVSITGVV-WODGFTADVT	298
QY	967	INTGSRATNGTVMASFGNGOTVNTYVATLTGASVATATNLNSNNVIOGOSTTFGFN	1026
DB	299	TMTGTAPYDQWQOLATLPDSGQRTINAMNASLTPSSGVTATGASHNARIAPGSLSFQ	358
QY	1027	GSYSGTNAAPT-----LSCT 1041	

OTHER INFORMATION: thermobacaeous
US-09-321-981-5

Query Match 11.5%; Score 637; DB 4; Length 386;
Best Local Similarity 40.6%; Pred. No. 1.9e-37;
Matches 149; Conservative 48; Mismatches 134; Indels 36; Gaps 15;

QY 698 VDAAGTASAPSTPVDCPTGPNON-----GVTSDGDEYRVQTNEMNSSAOQCLTINTATG 752
DB 31 VSTVIGTALALGPPAQA---NOQICDRYGTITID-RVYONNRNGTSATQC--INVGN 83
QY 753 AMVYTFANFSGTGAGAPATYPSIYKCHWNCCTTKNVGMPIQISQISAVTSKSTQVSS 812
DB 84 GFETIADGVSPTNGAPKSPSYVDGCHGNCAPRTT-LPMRISISGASPSVSRYXTGN 142
QY 813 GAYDAVDITNSTPTTGTGPNGTETIMILNSRGVQPFQSGQATGTVAGHTWVWQO 872
DB 143 GYVNAADIMIDPTPTPTNG-VNRTEIMTFNRVGPVOPISPVGT-AHVGRSMEVWTS 200
QY 873 QTSKTIISYVLTGATSIISNLDKAIADAAAGSLNTSD-YLIDVEAGEFIWQGGGLG 931
DB 201 NGSMDVLSF-LAPSAISWSFVDKD-FYDQAVSHGLATPDWYLTLSIQAGEFEWEGTGLA 258
QY 932 SNSFSVSVTSGG-----VACRATVYVNSDMGSGFTATVYVNTGSRATNGWTAVMS 982
DB 259 VNSFSANVAGNGGTGPTGPAACOVSYSTHT-WPGGFTVDTTINTGSTPVDGEMELDT 317
QY 983 FGNGOTVYNTYNTALTQSGASVTAIINLYSN-NVIOPGOSTTGFNGSYSGTN-AAPT--- 1037
DB 318 LPAHTVTSANNALISPASGAVTARSTGNSGRILANGTOSFGQTSSTGTFNAPAGGR 377
QY 1038 ---LSTCT 1041
DB 378 LNSTCT 384

RESULT 5
US-09-104-308-1
Sequence 1, Application US/09104308
Patent No. 6187577
GENERAL INFORMATION:
APPLICANT: Jones, Brian E.
APPLICANT: Van Der Kleij, Wilhelmus A.H.
APPLICANT: Weijer, Walter
TITLE OF INVENTION: No. 6187577e1 Cellulase Producing Actinomycetes,
TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSER: Genecor International, Inc.
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1013
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/104,308
FILING DATE: 24-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/974,042
FILING DATE: 19-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Stone, Christopher L.
REGISTRATION NUMBER: 35,696
REFERENCE/DOCKET NUMBER: GC539
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-846-7555

TELEFAX: 650-845-6504
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 371 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-104-308-1

Query Match 11.4%; Score 633.5; DB 4; Length 371;
Best Local Similarity 40.0%; Pred. No. 3.1e-37;
Matches 148; Conservative 52; Mismatches 131; Indels 39; Gaps 15;

QY 695 VAADAAGTASAPSTPVDCPTGPNON-----GVTSDGDEYRVQTNEMNSSAOQCLTINT 749
DB 16 LASTGALLTAAAPAO-----ANQICDRYGTITID-RVYONNRNGTSATQC--INV 65
QY 750 ATGAMTVSTANFSGTGAGAPATYPSIYKCHWNCCTTKNVGMPIQISQISAVTSKSTQ 809
DB 66 TGNGETIADGVSPTNGAPKSPSYVDGCHGNCAPRTT-LPMRISISGASPSVSRY 124
QY 810 VSSGAYDAVDITNSTPTTGTGPNGTETIMILNSRGVQPFQSGQATGTVAGHTWVW 869
DB 125 TGNVYNAADIMIDPTPTPTNG-VNRTEIMTFNRVGPVOPISPVGT-AHVGRSMEVW 182
QY 870 QCGQTSWKIISYVLTGATSIISNLDKAIADAAAGSLNTSD-YLIDVEAGEFIWQGG 928
DB 183 TGSNSNDVLSF-LAPSAISWSFVDKD-FYDQAVSHGLATPDWYLTLSIQAGEFEWEGT 240
QY 929 GIGNSNFSVSVTSGG-----VACRATVYVNSDMGSGFTATVYVNTGSRATNGWTAV 979
DB 241 GLAVNFSFSAVNAAGNGGTGPTGPAACOVSYSTHT-WPGGFTVDTTINTGSTPVDGEM 299
QY 980 AMSFGNGOTVYNTYNTALTQSGASVTAIINLYSN-NVIOPGOSTTGFNGSYSGTN-AAPT 1037
DB 300 DFTLPAGHTVTSVNNALISPASGAVTARSTGNSGRILANGTOSFGQTSSTGTFNAP 359
QY 1038 ---LSTCT 1041
DB 360 GARLNGTCT 369

RESULT 6
US-09-321-981-1
Sequence 1, Application US/09321981
Patent No. 6287839
GENERAL INFORMATION:
APPLICANT: Genecor International, Inc.
TITLE OF INVENTION: No. 6287839e1 Cellulase Producing Actinomycetes,
TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same
FILE REFERENCE: GC540-2
CURRENT APPLICATION NUMBER: US/09/321,981
CURRENT FILING DATE: 1999-05-28
PRIOR APPLICATION NUMBER: PCT/US99/11971
PRIOR FILING DATE: 1999-05-28
PRIOR APPLICATION NUMBER: 09/104,308
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 08/974,042
PRIOR FILING DATE: 1997-11-19
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 1
LENGTH: 371
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Nearest "neighbor" = Streptomyces
OTHER INFORMATION: thermobacaeous
US-09-321-981-1

Query Match 11.4%; Score 633.5; DB 4; Length 371;

```
Best Local Similarity 40.0%; Pred. No. 3.1e-37;
Matches 148; Conservative 52; Mismatches 131; Indels 39; Gaps 15;

QY 695 VAAVDAAGTASPTVPDCTPGPNON-----GVTSVODGEYRVQTNEMNSSAQCCTTTNT 749
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16 LASLIGALLTAAAPAO-----ANQOICDRYGTITID-RVYVQNNRMGTSAQC--INV 65

QY 750 ATGAATVSTANFSGGTGAPATPSIYKCHWNGCTTKNVGPIQISQIGSAVTSMTIQ 809
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 TGNGETIYQADGSVPTNAPKSPYSYDCHGNCAPRTT-LPMKISSIGSAPSSSVTRY 124

QY 810 VSSGAYDAVADYDWTNSTPTTGTGQPNGTETIMILNSRGVQPGPSOTATGTVAGHTWNV 869
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 125 TGNVYNNAYDIDMDPTPTNG-VNRTEIMIFNVRVGPQIGSPVGT-AHGGKSMWEW 182

QY 870 OGQOTSMKTIISVLTGATSIISNLDLKAIFADAAAGSLINTSD-YLLDVEAGFEIWOQO 928
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 TGSNSGNSNVISF-LAPSAISSWSFQVKD-FYDQAVSHGILATPDWYLTLSIQAGFEPEWEGT 240

QY 929 GLGSNSFSVTSVSG-----VACRATYYVNSDMGSGFATATYVNTGSRATNGWTV 979
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 241 GLAVNSFSAAVAGGNGGTGCTPCTPACOVSYSTHT-WPQGFVDTTITNTGSTPVDGMEL 299

QY 980 AMSFGNGOTVTNMTALTOSGASVATNLTYSN-NVIOPGOSTTGFNGSYSGTN-AAPT 1037
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 300 DFTLPAGHTVTSVWMLLSPASGAVTARSTGNSGRILANGTQSTGFCGTSSGAGFTAPA 359

QY 1038 -----LSCYT 1041
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 360 GARLNGTSCT 369

RESULT 7
US-09-216-295-21
; Sequence 21, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; TITLE OF INVENTION: No. 6268328el Variant EgIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216, 295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Actinomycece 11Ag8
US-09-216-295-21

Query Match 10.2%; Score 564.5; DB 4; Length 312;
Best Local Similarity 41.3%; Pred. No. 1.9e-32;
Matches 129; Conservative 43; Mismatches 109; Indels 31; Gaps 12;

QY 695 VAAVDAAGTASPTVPDCTPGPNON-----GVTSVODGEYRVQTNEMNSSAQCCTTTNT 749
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16 LASLIGALLTAAAPAO-----ANQOICDRYGTITID-RVYVQNNRMGTSAQC--INV 65

QY 750 ATGAATVSTANFSGGTGAPATPSIYKCHWNGCTTKNVGPIQISQIGSAVTSMTIQ 809
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 TGNGETIYQADGSVPTNAPKSPYSYDCHGNCAPRTT-LPMKISSIGSAPSSSVTRY 124

QY 810 VSSGAYDAVADYDWTNSTPTTGTGQPNGTETIMILNSRGVQPGPSOTATGTVAGHTWNV 869
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 125 TGNVYNNAYDIDMDPTPTNG-VNRTEIMIFNVRVGPQIGSPVGT-AHGGKSMWEW 182

QY 870 OGQOTSMKTIISVLTGATSIISNLDLKAIFADAAAGSLINTSD-YLLDVEAGFEIWOQO 928
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 TGSNSGNSNVISF-LAPSAISSWSFQVKD-FYDQAVSHGILATPDWYLTLSIQAGFEPEWEGT 240

QY 929 GLGSNSFSVTSVSG-----VACRATYYVNSDMGSGFATATYVNTGSRATNGWTV 979
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 241 GLAVNSFSAAVAGGNGGTGCTPCTPACOVSYSTHT-WPQGFVDTTITNTGSTPVDGMEL 299

QY 980 AMSFGNGOTVTNMTALTOSGASVATNLTYSN-NVIOPGOSTTGFNGSYSGTN-AAPT 1037
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 300 DFTLPAGHTVTSVWMLLSPASGAVTARSTGNSGRILANGTQSTGFCGTSSGAGFTAPA 359

QY 1038 -----LSCYT 1041
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 360 GARLNGTSCT 369
```

```
Db 241 GLAVNSFSAAVAGGNGGTGCTPCTPACOVSYSTHT-WPQGFVDTTITNTGSTPVDGMEL 299

QY 980 AMSFGNGOTVTN 991
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 300 DFTLPAGHTVTS 311

RESULT 8
US-09-118-319-9
; Sequence 9, Application US/09118319
; Patent No. 6114158
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Chen, Huizhong
; APPLICANT: Ljungdahl, Lars G.
; TITLE OF INVENTION: Orlinomyces Cellulase Celf Protein and Coding Sequences
; FILE REFERENCE: 33-98sequence listing
; CURRENT APPLICATION NUMBER: US/09/118, 319
; CURRENT FILING DATE: 1998-07-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 461
; TYPE: PRT
; ORGANISM: Fusarium oxysporum
US-09-118-319-9

Query Match 10.1%; Score 556.5; DB 3; Length 461;
Best Local Similarity 32.3%; Pred. No. 1.3e-31;
Matches 149; Conservative 71; Mismatches 167; Indels 75; Gaps 16;

QY 17 TTRMPAIRKRLAGVLAGAVSIASIVPLAMQHPALATHVDNPYAGATFFVNPYMAQEV 76
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 70 TAAGPSSTATKTATNGSSSTTAGSVSA---PPAAS---DNFAYGVDIANNNTYRSEV 123

QY 77 OSEA-----NQTATLAAKRVSTYSTAVWMDRIALINGVNGGPELTYLLDAALSQOQGT 133
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 124 NMLAVPKLSGAKATAAAYADVPSFO---WMDTYDHISLME---DTLADIRKANKAG- 174

QY 134 PREVEIETIYDLPGRDCAALASNGELPATAAGLQTYENQYIDPIASITSNPYSSLRIYV 193
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 175 GKTYAQFVYTDLPNRDCAAASNGEYSLDKDQANKYKA-YINKIGILION--YSTQKYL 231

QY 194 IIEPDSLPAVNTNMSIOACATAPVPEYEGIEYALTKLHAIPNVYLYMDAASHGWLGMVNN 253
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 VIEPDSLNLVNTLNVWDCAKAESAYKELTYVYAIKELN-LPVSVMYLDAGHGMWGMFAN 290

QY 254 ASGYQEVQKVLNASIGVNGIDGFTYNTANTYPLKEPKMTATQOYGQPVESANFYQNNP 313
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 291 IGPAAKLYAQIYKDAKPSRVGLVTNVSNN-----GWLKSTKPDYTESNP 337

QY 314 DIDEADVAVDLTSRLVYAAGFPSSIGMLIDTLNMGWGNGEPETATDVNTFVQSKID 373
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 338 NYDEQRT-IMAFAPLLADQSGNSVKEFYDQGRSG---KQPLG----- 375

QY 374 LRQHRGLMGNQNGAGIGOPPOASPTDFPNAHLDAVWIKPPESEDSGTSAAASPTTGKESD 433
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 376 -QKAGDMCNMAKGTGFLRPSTNTGD--ALADATVWVKPGGESGJTS---DTSAAYTD 427

QY 434 PYCDDPTYYTSYGLTNALPNSPIAGQWPPAOPDOLVANAARA 475
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 428 YHCG-----LDALKPAPEAGCTWFOAYFEQLLDNANPS 459
```

```
RESULT 9
US-09-286-691-25
; Sequence 25, Application US/09286691
; Patent No. 6190189
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
```



```
Db 47 SSGTCVSNDDYYSCQLPGAASSSSSTRASTTRVSPPTSSSSATPPPG--STTTAVPV 104
QY 46 AMOHPAIAATHVNDNPYAGATFFVNPYMAOEVOSEANQTNATLAAMKRVSTYSTAVMD 105
Db 105 G-----SGTATYSGNPFVGTYPMANAYASEVSSLAIPSLTGAMATAAAVAKVSPFMD 160
QY 106 -----RTAINGVNGGPGLTYYLDALSSQOQGTPEVIEIYIYDLPGRDCAALAS 155
Db 161 TLDKTPPLMEQTLADIRTKANGG--NYAG-----QFVYVDLPDRDCAALAS 204
QY 156 NCELPAATAAGTQYETQYIDPASIILSNPKYSLSRIYVITIEPDSLPNATVNNISIQACATA 215
Db 205 NCEYSIADGSAKK--NYIDTIRQIV--VEYSDIRTLVIEPDSLANTVNLGTPRCANA 261
QY 216 VPEYEGIEYALTKLHAIDPNVITYMDAHSGLWGPNNASGVYQEVOKYLNASIGVNGID 275
Db 262 OSATYECIYIAVYQOLN-LPNVAMYIDAGHAGWLGMPANDPPAAQLFANVYKNASSPRALR 320
QY 276 GFVTNANTPLKEPMTATQOYGQPVESANFYQNNPDIDEADYAVDLXSLVAAFGFS 335
Db 321 GLATNVAANTN-----GWNITSPSTQGNNAVYNEKLYTHAIGPLLANHGW-S 366
QY 336 SIGMLIDTLRNGMGNEPTGSTATDVNTFVNSKIDLRHRLGMLCNOAGAGLQGPQA 395
Db 367 NAFETIDGSRG---KQPTG-----QQOKGDMCNVIGTGFGRPSA 404
QY 396 SPDPFPNAHLDAVYWKIPGESDGTSAASDPPTGKKSDPMCDPTTYTSYGVLTNALPNSP 455
Db 405 NMGD---SLIDSFVWVKPGCEDGTSSSAP---RFDSHC-----ALPDALQAP 448
QY 456 IAGOMFPFAOFDOLVANARPA 475
Db 449 QAGAMFOATFVOLLTNANPS 468

RESULT 12
US-09-142-759-1
; Sequence 1, Application US/09142759A
; Patent No. 6127160
; GENERAL INFORMATION:
; APPLICANT: YAMANOBE, Takashi
; APPLICANT: WATANABE, Manabu
; APPLICANT: HAMAYA, Toru
; APPLICANT: SUMIDA, Naomi
; APPLICANT: AOYAGI, Kaoru
; APPLICANT: MORAKAMI, Takeshi
; TITLE OF INVENTION: PROTEIN HAVING CELLULASE ACTIVITY AND PROCESS FOR
; FILE REFERENCE: 051673
; CURRENT APPLICATION NUMBER: US/09/142,759A
; EARLIER FILING DATE: 1998-09-14
; EARLIER APPLICATION NUMBER: PCT/JP97/00824
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: P. HEI-8-84479
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 457
; TYPE: PRT
; ORGANISM: Acromonium cellulosilyticus
US-09-142-759-1
```

Query Match 9.8%; Score 544; DB 3; Length 457;

Best Local Similarity 32.5%; Pred. No. 9.5e-31; Mismatches 171; Indels 82; Gaps 16;

```
QY 16 GTTRPAIRKRLRAGVLAASIVPLAQHPAIAATHVNDNPYAGATFFVNPYMAOE 75
Db 74 GTT-SPPTTTTAKSTATTATAASG-----NPSGYQLVNPYSSB 115
QY 76 VOSEANQNTNATLAAMKRVSTYSTAVMMDRITAINGVNGGPGLTYYL-DALSSQOQGT 134
```

```
Db 116 VHTLAIPSLTGLAAATRAAETIPSFVWLDTAKY-----PTMGTYLANIEANRAGAS 169
QY 135 PEVIEI-VIYDLPGRDCAALASNGELPATAGIOTETQYIDPASIILSNPKYSLSRIY 193
Db 170 PPIAGIFVYDLPDRCAALASNGELTAVANGYANKA-YIDSIVAQLR--AYPDVHTL 226
QY 194 IIEPDSLPAVNTNMSIQACATAVPEYEGIEYALTKLHAIDPNVITYMDAHSGLWGPNN 253
Db 227 IIEPDSLAMNVTNLSAKCAEAQASAYEECVNVALNLN-LANVAMYIDAGHAGWLGWSAN 285
QY 254 ASGYVOEKVLNASIGVNGIDGFYNTANTYTPKPEPMTATQOYGQGVQPVESANFYQNN 313
Db 286 LSPAAQLFATVYKNASAPASLGLATNVANYN-----AMSSISPP-----SYTSGDS 332
QY 314 DIDEADYAVDLXSLVAAAFPSISGMLIDTLRNGMGNEPTGSTATDVNTFVNSKID 373
Db 333 NYDEKILYIALSPILTSKNMPNA-HFIMDTSRNG---VQPT----- 369
QY 374 LKQHRGLCNOAGAGLQGPQASPTDFPNAHLDAVYWKIPGESDGTSAASDPPTGKKSD 433
Db 370 KOQAMGDMCNVIGTGFGRVQPTNTGQD---PLEDAFVWVKPGCEDGTSSS---ATRYD 422
QY 434 PMCDPTTYTSYGVLTNALPNSPFAGOMFPFAOFDOLVANARPAVY 477
Db 423 FHCG-----YSDALQAPAPAGTWFQATFVOLLTNANPALV 457
```

RESULT 13
US-09-286-691-24

```
; Sequence 24, Application US/09286691
; Patent No. 6190189
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96
; CURRENT APPLICATION NUMBER: US/09/286,691
; EARLIER FILING DATE: 1999-04-05
; EARLIER APPLICATION NUMBER: US 60/027,883
; EARLIER FILING DATE: 1996-10-04
; EARLIER APPLICATION NUMBER: PCT US97/18008
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-286-691-24
```

Query Match 9.7%; Score 538; DB 4; Length 360;

Best Local Similarity 33.5%; Pred. No. 1.8e-30; Mismatches 143; Conservative 50; Mismatches 152; Indels 82; Gaps 13;

```
QY 59 NPYAGATFFVNPYMAOEVOSEANQTNATLAAMKRVSTYSTAVVMD-----RTA 108
Db 4 NPEVGTYPMANAYASEVSSLAIPSLTGAMATAAAVAFSPFMDITDTPMLMEQTLA 63
QY 109 AINGVNGGPGLTYYLDALSSQOQGTPEVIEIYIYDLPGRDCAALASNGELPATAGIOT 168
Db 64 DIRTKANGG--NYAG-----QFVYVDLPDRDCAALASNGEYSIADGSAKK 107
QY 169 YENOYDPLASILSNPKYSLSRIYVITIEPDSLPNATVNNISIQACATAVPEYEGIEYAL 228
Db 108 YK-NYIDTIRQIV--VEYSDIRTLVIEPDSLANTVNLGTPRCANAQASAYIECTINAVT 164
QY 229 KLHAIDPNVITYMDAHSGLWGPNNASGVYQEVOKYLNASIGVNGIDGFVTNTANTYTPK 288
Db 165 QLN-LPNVAMYIDAGHAGWLGMPANDPPAAQLFANVYKNASSPRALRGLATNVANYN--- 220
```

```
QY 289 EPFMTATQVCGPVESEANFYQWNPDIIDEADYAVDLSRLVAAGFPSSIGMLIDTLRWGM 348
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 221 -----GWNITSPSPYQGNAYVNEKLYIHAIQPLANHGW-SNAFFITDQGRSG- 268
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 349 GGPNEPTGPSTATDVNTFVNQSKIDLRHRLGMCNONGAGLGOPPOASPTDFPNAHLDAY 408
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 269 ---KQPTG-----OQMGDKMKNVICTGFGIRPSANTGD---SLDSF 304
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 409 VWKPGESDGTSAASDPTTGKKSDDPMCDPTTYSYGLVTALNPSPNATGMPAOPDOL 468
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 305 VWKPGEGCDGTSDSSAP-----RFDSHC-----ALPDALQPPAQAGAMFOAYFYOL 351
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 469 VANARPA 475
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 352 LTNNMPS 358
```

```
RESULT 14
US-09-687-147-24
; Sequence 24, Application US/09687147
; Patent No. 6268198
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96a
; CURRENT APPLICATION NUMBER: US/09/687,147
; PRIOR FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US 60/027,883
; PRIOR FILING DATE: 1996-10-04
; PRIOR APPLICATION NUMBER: PCT US97/18008
; PRIOR FILING DATE: 1997-10-03
; PRIOR APPLICATION NUMBER: 09/286,691
; PRIOR FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 24
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-687-147-24
```

```
Query Match 9.7%; Score 538; DB 4; Length 360;
Best Local Similarity 33.5%; Pred. No. 1.8e-30;
Matches 143; Conservative 50; Mismatches 152; Indels 82; Gaps 13;
```

```
QY 59 NPVAGATFEVNPVMAQVQSEANQNTAATLAAMKRVYSTYSTAVWMD-----RIA 108
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 4 NPEVGVPWMAAYASEVSSLAIPSLTGAMATAAAVAKVPSFWMLDTLDRKPLMEQTILA 63
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 109 AINGVNGGPGITTYLDAALSOOGCTTPEVIEIYIDLPGRCALASNGELPATAGLOT 168
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 64 DIRANNNGG--NYAG-----QFVYIDLPRDCAALASNGEYSIADGVAK 107
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 169 YETQYIDPLASILSNPKYSSLRITVITIEPDSLPAVNTMSTIQACATVPVYEGGIEYALT 228
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 108 YK-NYIDTIRQIV--VEYSDIRTLVIEPDSLANTLNTLGPCKANQASALIECINAVT 164
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 229 KLAHAIPIVYIYMDAHSGLMGWPNNASGYVOEVOKVLNASIGVNGIDGFTVNTANTPLK 288
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 165 QLN-LPNNVAMYLDAHGAGLGMWPANODPAAQLFANVYKNASSPRALRLGLATNVAN 220
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 289 EPFMTATQVCGPVESEANFYQWNPDIIDEADYAVDLSRLVAAGFPSSIGMLIDTLRWGM 348
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 221 -----GWNITSPSPYQGNAYVNEKLYIHAIQPLANHGW-SNAFFITDQGRSG- 268
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 349 GGPNEPTGPSTATDVNTFVNQSKIDLRHRLGMCNONGAGLGOPPOASPTDFPNAHLDAY 408
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 269 ---KQPTG-----OQMGDKMKNVICTGFGIRPSANTGD---SLDSF 304
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 409 VWKPGESDGTSAASDPTTGKKSDDPMCDPTTYSYGLVTALNPSPNATGMPAOPDOL 468
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
```

```
DB 305 VWKPGEGCDGTSDSSAP-----RFDSHC-----ALPDALQPPAQAGAMFOAYFYOL 351
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 469 VANARPA 475
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 352 LTNNMPS 358
```

```
RESULT 15
US-08-169-948B-12
; Sequence 12, Application US/08169948B
; Patent No. 5861271
; GENERAL INFORMATION:
; APPLICANT: Fowler, Timothy
; APPLICANT: Ward, Michael
; APPLICANT: Clarkson, Kathleen
; APPLICANT: Collier, Katherine
; APPLICANT: Larens, Edmund
; TITLE OF INVENTION: No. 5861271el Cellulase Enzymes and Systems
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 180 Kimball Way
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk-
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/169,948B
; FILING DATE: DEC 17 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Horn, Margaret A.
; REGISTRATION NUMBER: 33,401
; REFERENCE/DOCKET NUMBER: GC226
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7536
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 365 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-169-948B-12
```

```
Query Match 9.7%; Score 538; DB 2; Length 365;
Best Local Similarity 33.3%; Pred. No. 1.8e-30;
Matches 144; Conservative 51; Mismatches 155; Indels 82; Gaps 13;
```

```
QY 54 ATHVNPVAGATFEVNPVMAQVQSEANQNTAATLAAMKRVYSTYSTAVWMD----- 105
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 4 ATISGNPFVGVTPWMAAYASEVSSLAIPSLTGAMATAAAVAKVPSFWMLDTLDRKPLM 63
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 106 --RIAINGVNGGPGITTYLDAALSOOGCTTPEVIEIYIDLPGRCALASNGELPAT 163
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 64 EGTLDITRANKNG--NYAG-----QFVYIDLPRDCAALASNGEYSIND 107
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 164 AGLOTYETQYIDPLASILSNPKYSSLRITVITIEPDSLPAVNTMSTIQACATVPVYEGGI 223
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 108 GGVARYK-NYIDTIRQIV--VEYSDIRTLVIEPDSLANTLNTLGPCKANQASALIECI 164
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 224 EYALFKLAHAIPIVYIYMDAHSGLMGWPNNASGYVOEVOKVLNASIGVNGIDGFTVNTAN 283
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 165 NYAVTQLN-LPNNVAMYLDAHGAGLGMWPANODPAAQLFANVYKNASSPRALRLGLATNVAN 223
      |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
```

```

Oy 284 YTPLEPFMTATQYGGOPVESANFYQWNPDIADYAVDLYSRLYAGPSSIGMLIDT 343
Db 224 YN-----GWNITSPSTYQGNAYTNEKLYIHAIGPLANNGM-SNAFFITDQ 269
Oy 344 LKNGWGNPEPTGPSTATDVNTFVNQSKIDLROHRLMCNONGAGIGPPQASPTDFPNA 403
Db 270 GRS-----KQPTG-----QQMGDMCNVIGTGFIRPSANTGD---S 304
Oy 404 HLDATVWIKPPGESDGTSAADPTTGKKSDPMCDPTTYSYGLTNALPNSPIAGOWPFA 463
Db 305 LLDSEVWVKPGGECDDGTSDSSAP---RDSHC-----ALPDALQAPAPQAGAMFOA 351
Oy 464 QFDOLVANARPA 475
Db 352 YFVQLITNANPS 363

```

Search completed: August 29, 2002, 16:22:37
 Job time: 367 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:21:46 ; Search time 127 Seconds
(without alignments)
202.032 Million cell updates/sec

Title: US-09-917-384-7
Perfect score: 1244

Sequence: 1 DCTPCPNQNGVTYSVDGEYR.....IMGGGLGNSNSFSVTS 231

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq.032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	472.5	38.0	371	20	AAV06367
2	472.5	38.0	371	20	AAV08473
3	472.5	38.0	371	21	AAV14880
4	472.5	38.0	371	21	AAV84345
5	472.5	38.0	371	21	AAV67496
6	472.5	38.0	386	21	AAV67497
7	472.5	34.1	381	20	AAV06368
8	424.5	34.1	381	21	AAV14881
9	409	32.9	260	21	AAV84346
10	407.5	32.8	260	20	AAV06369
11	407.5	32.8	260	21	AAV14882

12	407.5	32.8	260	21	AAV84347
13	260.5	20.9	259	20	AAV06351
14	260.5	20.9	259	21	AAV14864
15	260.5	20.9	259	21	AAV84329
16	256	20.6	104	20	AAV06337
17	236.5	19.0	348	20	AAV06362
18	236.5	19.0	348	21	AAV14875
19	236.5	19.0	348	21	AAV84340
20	221.5	17.8	250	22	AAU00230
21	218.5	17.6	261	20	AAV88462
22	215	17.3	104	20	AAV06343
23	215	17.3	264	20	AAV06370
24	215	17.3	264	21	AAV14883
25	215	17.3	264	21	AAV84348
26	208.5	16.8	246	20	AAV06366
27	208.5	16.8	246	21	AAV14879
28	208.5	16.8	246	21	AAV84344
29	206.5	16.6	238	20	AAV06358
30	206.5	16.6	238	21	AAV14871
31	206.5	16.6	238	21	AAV84336
32	204	16.4	239	20	AAV06353
33	204	16.4	239	21	AAV14866
34	204	16.4	239	21	AAV84331
35	192.5	15.5	239	18	AAV14595
36	192.5	15.5	239	18	AAV15562
37	192	15.4	247	20	AAV14870
38	192	15.4	247	21	AAV84335
39	192	15.4	247	21	AAV84335
40	190.5	15.3	244	19	AAV68593
41	190.5	15.3	244	21	AAV44341
42	188.5	15.2	237	20	AAV06364
43	188.5	15.2	237	21	AAV14877
44	188.5	15.2	237	21	AAV84342
45	187	15.0	238	20	AAV06361

ALIGNMENTS

RESULT 1
AAV06367 standard; Protein; 371 AA.
XX AAV06367;
XX 06-SEP-1999 (first entry)
XX Streptomyces sp. EGIIT-like cellulase.
XX Cellulase; endoglucanase; EGIIT; textile; feed additive; baking;
XX food processing; grain wet milling; pulp; paper.
XX Streptomyces sp.
XX W09931255-A2.
XX 24-JUN-1999.
XX 14-DEC-1998; 98WO-US26552.
XX 16-DEC-1997; 97US-0991720.
XX (GENV) GENENCOR INT INC.
XX Bower BS, Fowler T, Phillips JT;
XX WPI; 1999-395187/33.
XX EGIIT like cellulase
XX Example; Fig 6; 47pp; English.
XX The present polypeptide represents a full-length sequence of a

Amino acid sequenc
Aspergillus aculea
Aspergillus aculea
Amino acid sequenc
Streptomyces sp. E
Glicolactium roseum
Glicolactium roseum
Amino acid sequenc
M. cinnamomea xylo
Streptomyces livid
Erythrina carotovora
Erythrina carotovora
Amino acid sequenc
Emeticella deserto
Actinomyces 11A68
Amino acid sequenc
Fusarium equiseti
Fusarium equiseti
Aspergillus kawach
Aspergillus kawach
Amino acid sequenc
Aspergillus niger
Cellulase polypt
Chaetomium brasill
Chaetomium brasill
Amino acid sequenc
Tiarosporella phas
T. phaseolina xylo
Glicolactium roseum
Glicolactium roseum
Amino acid sequenc
Glicolactium roseum

XX Novel variant endoglucanase III-like cellulases with improved
PT surfactant stability and resistance to temperature stress, useful for
PT textile processing or cleaning, treating wood pulp, food and grain, and
PS reducing biomass to glucose
XX

Disclosure: Page 65-66; 73pp; English.

CC The present sequence represents an endoglucanase III (EGIII)-like
CC cellulase. The cellulase has homology to the Trichoderma reesei EGIII
CC protein. The variant cellulases have improved temperature stability,
CC and improved surfactant stability. The variant cellulases and
CC compositions containing them are used in textile processing or cleaning,
CC e.g. stonewashing of indigo dyed denim, and modifying the texture, feel
CC or appearance of cellulose containing fabrics (e.g. improving fabric
CC smoothness or removing pills and fibrils). The compositions may also be
CC used for the removal of immature or dead cotton from cellulosic fibres
CC or fabric, which can cause uneven dyeing. The cellulase may also be used
CC in a detergent composition for washing laundry and dishes and in the
CC treatment of animal feed, wood pulp, paper, non-animal foods and grains.
CC The enzymes may also be used in the reduction of biomass to glucose.
XX

SQ Sequence 429 AA:

Query Match 32.9%; Score 409; DB 21; Length 429;
Best Local Similarity 41.6%; Pred. No. 1.8e-27;
Matches 91; Conservative 32; Mismatches 90; Indels 6; Gaps 6;

OY 13 SVQGEYRVQTEENWSSAOCLTINTAGMTVSTANFSGTGAPATYPSIKGCHWGN 72
DB 99 svfngghyqnmrwgstapcvta-tdtg-frvtgdagapnbgakpsvrvngchyn 156
OY 73 CTTKNVGMPIQISQISAVTSMSTVOVSSGADVAVDIWNTSTPTTGQPNGTIMWLN 132
DB 157 c-spgrdlpvridtvaapsisysgfvdgavyvnaasydiwldpcartdg-vnqetmlwfn 214
OY 133 SRGVOVPSQFATGVTVAGHTVMWVMOGQOTSMKIISYVLTPEATGISNLDKAIKTRADA 192
DB 215 rvgpdpdpispygt-asvvggrtewvsgngsgndvlsiv-apsalsqwsldvmdfvraty 272
OY 193 ARGSLNTSDYLLDVEAGFEIMOGGOGIGSNSFSVSYSVTS 231
DB 273 arglaendwyltsvqgafepwqngaglavnsfstvetytg 311

RESULT 10

AA06369
ID AAY06369 standard; Protein; 260 AA.

XX AAY06369;

DT 06-SEP-1999 (first entry)

DE Rhodothermus marinus EGIII-like cellulase.

KW Cellulase; endoglucanase; EGIII; textile; feed additive; baking;
KW food processing; grain wet milling; pulp; paper.

OS Rhodothermus marinus.

PN WO9931255-A2.

PD 24-JUN-1999.

PF 14-DEC-1998; 98WO-US26552.

PR 16-DEC-1997; 97US-0991720.

PA (GENV) GENENCOR INT INC.

XX Bower BS, Fowler T, Phillips JT;

DR WPI; 1999-395187/33.

XX EGIII like cellulase

XX Example: Fig 6; 47pp; English.

CC The present polypeptide represents a full-length sequence of a
CC novel EGIII-like cellulase of Rhodothermus marinus. It was
CC deduced from a gene sequence isolated from genomic DNA using PCR
CC primers (see AAX59180-91) based on conserved motifs (see AAY06325-29)
CC of Trichoderma reesei EGIII cellulase and related enzymes. PCR
CC has been used to identify novel EGIII-like enzymes, including the
CC present protein, from bacterial and fungal sources (see AAY06331-70).
CC The sequence shows homology to T. reesei EGIII (see AAY06330). Also
CC provided by the invention are vectors, host cells and methods
CC for the recombinant production of such enzymes, which can be used
CC in the treatment of cellulose-containing textiles, as feed
CC additives, in the treatment of wood pulp, in the reduction of
CC biomass to glucose, in the treatment of indigo dyed denim, or
CC as laundry detergent components (all claimed).

SQ Sequence 260 AA:

Query Match 32.8%; Score 407.5; DB 20; Length 260;
Best Local Similarity 40.1%; Pred. No. 1.3e-27;
Matches 89; Conservative 30; Mismatches 84; Indels 19; Gaps 7;

OY 14 VQGEYRVQTEENWSSAOCLTINTAGMTVSTANFSGTGAPATYPSIKGCHWGN-- 71
DB 50 vaggrrvrvlnvwaagtaqlevgletgnfticrhadhdgn--vaaypalyfychwapa 107
OY 72 ----NCTKNVGMPIQISQISAVTSMSTVOVSSGADVAVDIWNTS-TPTTGQPNGTIE 126
DB 108 rairdcaar-agaavrarahel-----dvrlptlgrwnaaydiwfsfpvtnsgnygsgeae 159
OY 127 IMTWLNSRGVOPFGSOTANGVYAGHTVMWVMOGQOTSMKIISYVLTPEATGISNLDKIA 186
DB 160 lmlwlnngvgmpgsarval-velagatwewwa-dwdwnylayrrtptltsvseidlka 217
OY 187 IFADAAARGSLNTSDYLLDVEAGFEIMOGGOGIGSNSFSVS 228
DB 218 flddavaraylrpewylhavevtfelwegsgaglrtdafstvtv 259

RESULT 11

AAB1482
ID AAB1482 standard; Protein; 260 AA.

XX AAB1482;

DT 21-NOV-2000 (first entry)

DE Emeritella desertoru EGIII-like cellulase.

KW Emeritella desertoru; Trichoderma reesei; endoglucanase III; EGIII;
KW cellulase; mutant; enzyme stability; textile treatment;
KW wood pulp treatment; feed additive; detergent.

OS Emeritella desertoru.

PN WO200037614-A2.

PD 29-JUN-2000.

PF 12-NOV-1999; 99WO-US26704.

PR 18-DEC-1998; 98US-0216295.

PA (GENV) GENENCOR INT INC.

XX Mitchinson C, Wendt DJ;


```

XX WP1; 1999-395187/73.
XX
XX EGIIT like cellulase
XX
XX Example; Fig 6; 47pp; English.
XX
CC The present polypeptide represents a full-length sequence of a
CC novel EGIIT-like cellulase of Aspergillus aculeatus. It was
CC deduced from a gene sequence isolated from genomic DNA using PCR
CC primers (see AA59180-91) based on conserved motifs (see AA706325-29)
CC of Trichoderma reesei EGIIT cellulase and related enzymes. PCR
CC has been used to identify novel EGIIT-like enzymes, including the
CC present protein, from bacterial and fungal sources (see AA706331-70).
CC The sequence shows homology to T. reesei EGIIT (see AA706330). Also
CC provided by the invention are vectors, host cells and methods
CC for the recombinant production of such enzymes, which can be used
CC in the treatment of cellulose-containing textiles, as feed
CC additives, in the treatment of wood pulp, in the reduction of
CC biomass to glucose, in the stone washing of indigo dyed denim, or
CC as laundry detergent components (all claimed).
XX
SQ Sequence 259 AA;
XX
Query Match 20.9%; Score 260.5; DB 20; Length 259;
Best Local Similarity 33.9%; Pred. No. 8,4e-15;
Matches 82; Conservative 21; Mismatches 96; Indels 43; Gaps 10.
OY 17 GERYQTNEMN---SSAQCLTINTATGAWY-STA-NFSGTGAPATYPSIYKCHWG 71
DB 31 gylltnmlwlgwdagsgsgctctvnassasgftswtkmwsjggenyvs-----yan 81
OY 72 NCTTKNMGPIQISOIGSAVTSWSTQVSSGAYDAVDIMTNSPTTGGQNGTEIMWL 131
DB 82 sgltfnkkkivsgsqf-pttarwsydnbgira-dvayaldftaadlnhvtwsgdyelmvl 139
OY 132 NSRGVOPFGSGTATGCVYAGHTNWNWQOQTSKWKIISVLTLPGATSISNLDKAIPLDA 191
DB 140 arygvgvqpsgsqfat-atvdgqltwelwygangsqtksfivapdtltsfsg-dvndfky1 197
OY 192 AA-RSLTMSDYILDV-----EAGFETMOGGOGIGSNSFSV 226
DB 198 tqnhfpassqyltlcqlgftpeptfggpattlsvnsasvsgagfepwngaglaavnstis 257
OY 227 SV 228
DB 258 tv 259
XX
RESULT 14
AABI4864
ID AABI4864 standard; Protein; 259 AA.
XX
AC AABI4864;
XX
DE 21-NOV-2000 (first entry)
XX
XX Aspergillus aculeatus EGIIT-like cellulase.
XX
KW Aspergillus aculeatus; Trichoderma reesei; endoglucanase III; EGIIT;
KW cellulase; mutant; enzyme stability; textile treatment;
XX wood pulp treatment; feed additive; detergent.
XX
OS Aspergillus aculeatus.
XX
PN WO200037614-A2.
XX
PD 29-JUN-2000.
XX
PF 12-NOV-1999; 99WO-US26704.
XX
PR 18-DEC-1998; 98US-0216295.

```

XX	(GENM) GENENCOR INT INC.
PA	Mitchinson C, Wendt DJ;
XX	WPI; 2000-482483/42.
DR	
XX	Novel endoglucanase III or endoglucanase III-like cellulase useful for
PT	treating textiles and wood pulp compris a substitution or deletion at
PT	specified positions in the wild form of endoglucanase III -
XX	
PS	Example 1; Fig 3; 52pp; English.
XX	
CC	The present sequence is a cellulase related to endoglucanase III (EGIII)
CC	from Trichoderma reesei. EGIII-like genes were isolated from genomic DNA
CC	libraries constructed from various microorganisms by PCR. The isolated
CC	genes showed significant homology to EGIII from T. reesei. Certain
CC	substitution and deletion mutations have been incorporated into EGIII and
CC	EGIII-like cellulases to produce variant enzymes with improved stability,
CC	e.g. increased resistance to temperature stress. The mutants may be used
CC	in textile and wood pulp treatment, as a feed additive, and for reducing
CC	biomass to glucose. They are also useful for stonewashing or indigo dyed
CC	denim and as an agent in laundry and dish detergents.
SO	Sequence 259 AA;
Query Match	20.9%; Score 260.5; DB 21; Length 259;
Best Local Similarity	33.9%; Pred.No. 8.4e-15;
Matches	82; Conservative 21; Mismatches 96; Indels 43; Gaps 10;
OY	17 GEYRQQTENWN-----SSAQCCTINTATGAMV-STN-NFSGGTGAPATYPSTIYGCHWG 71
Dd	: : : : : : : : : : : : : : : : : :
OY	31 gyltlmnlwqkdagsqsqtvltnsaasagtslkwmsgenavks-----yan 81
OY	72 NCTTKNVCMPLIOISIGSAVTSMSTRTOYSSGAYDVADYDWTNSTPRTTCQPNGTEIMVL 131
Dd	: : : : : : : : : : : : : : : : :
OY	82 sgltknkkrivsdsj-ptlarwsydnctgir-dvaydlftadinhvtwsgdyelmwl 139
Dd	: : : : : : : : : : :
OY	132 NSRGGVCFPGSQTAGTVVAGHTMMVMWOQQTSMKRIISYLTPCATSISSLDKAIFFADA 191
Dd	
OY	140 arysgvqlgsgrat-atvdggtwelwygangsqrktystfvaptpitlsifdg-dvndtfkl 197
OY	192 AAKSLMTSDYLDLV-----EAGFEITMOGGOIGSNSFSV 226
Dd	: : : : : : : : : : :
OY	198 tqnhgfpassgyllltlgfgteptfgspatlsvnsasvyvgagfepwpqaglavnfsf 257
OY	227 SV 228
Dd	: :
OY	258 tv 259
RESULT 15	
ID	AAY84329
ID	AAY84329 standard; Protein; 259 AA.
XX	
AA	AAY84329;
AC	
DE	12-JUL-2000 (first entry)
XX	
AM	Amino acid sequence of an endogluacanase III (EGIII)-like cellulase.
XX	
END	Endoglucanase III; EGIIL; EGIIR-like cellulase; surfactant stabilliy;
KW	cellulase; textile processing; textile cleaning; stonewashing;
KW	indigo dyed denim; cellulose containing fabric; fabric smoothness;
KW	pill removal; fibril removal; cotton; celulosic fibre; dyeing; detergent;
KX	animal feed; wood pulp; paper; grain; biomass reduction; glucose.
OS	Aspergillus aculeatus.
NN	WO200014208-A1.
DD	16-MAR-2000.

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:24:14 ; Search time 72.78 Seconds
(without alignments)
304.983 Million cell updates/sec

Title: US-09-917-384-7
Perfect score: 1244
Sequence: 1 DCTPGPNQNGVTSVQDEYR.....IMOGGGLGNSFSVSVTSG 231

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*

1: p1r1.*
2: p1r2.*
3: p1r3.*
4: p1r4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	436.5	35.1	382	JC2571	cellulase (EC 3.2.
2	235.5	18.9	151	H70895	hypothetical prote
3	234	18.8	237	S12610	cellulase (EC 3.2.
4	215	17.3	264	JU0328	cellulase (EC 3.2.
5	184.5	14.8	239	S55931	cellulase (EC 3.2.
6	133	10.7	334	G90360	endoglucanase prec
7	139	10.4	1286	S28634	adhesin AIDA-I pre
8	138	10.3	332	G90291	endoglucanase
9	112	9.0	1025	G81722	polymorphic membra
10	108	8.7	1275	T33369	hypothetical prote
11	108	8.7	1777	T34360	hypothetical prote
12	105.5	8.5	258	H72240	endoglucanase - Th
13	105.5	8.5	958	A82583	conserved hypothet
14	105	8.4	1036	A29832	HPI layer surface
15	104	8.4	747	B47093	cellulase (EC 3.2.
16	100	8.0	4776	E95206	cell wall surface
17	98.5	7.9	526	A56573	nuclear pore compl
18	98	7.9	1390	T18883	hypothetical prote
19	98	7.9	1749	S75138	hypothetical prote
20	98	7.9	3075	S14458	laminin alpha-1 ch
21	97.5	7.8	562	AE1335	probable peptidogl
22	97.5	7.8	5188	B85547	probable RTX fami
23	97.5	7.8	5291	E90696	hypothetical prote
24	97.5	7.8	13286	T03099	mucin, submaxillar
25	97	7.8	989	D89852	fibrogen-binding
26	96.5	7.8	948	C75265	hexagonally packed
27	96.5	7.8	1004	T00046	surface layer prot
28	96.5	7.8	1076	A35622	nuclear pore prote
29	96.5	7.8	1649	C86822	hypothetical prote

30	96.5	7.8	3716	2	E70969	probable ppe prote
31	96	7.7	479	2	C90457	cytochrome b558/56
32	96	7.7	1441	2	A86685	prophage p11 prote
33	96	7.7	1475	2	S42718	nuclear pore compl
34	95.5	7.7	839	2	H97758	outer membrane ass
35	95	7.6	211	2	C37491	hypothetical prote
36	95	7.6	221	2	UC7307	endo-1,4-beta-xy1a
37	95	7.6	666	2	A42296	lysosyme 2 (EC 3.2
38	95	7.6	1345	2	H90975	hypothetical prote
39	95	7.6	1461	2	E90696	hypothetical prote
40	95	7.6	1461	2	A85547	hypothetical prote
41	95	7.6	2232	2	T34434	probable invasin 2
42	95	7.6	2660	2	E85822	hypothetical prote
43	94.5	7.6	1251	2	T21389	immunodominant ant
44	94	7.6	233	2	B90062	hypothetical prote
45	94	7.6	657	2	S25184	csp1 protein - Cor

ALIGNMENTS

RESULT 1
JC2571
cellulase (EC 3.2.1.4) precursor - Streptomyces rochei (strain A2)
N:Alternate names: endo-1,4-beta-glucanase; endoglucanase
C:Species: Streptomyces rochei
C:Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 22-Oct-1999
C:Accession: JC2571; S34392
R:Porto, B.; Hanhart, E.; Irdani, T.; Iqbal, M.; McCarthy, A.J.; Mastrometi, G.
Gene 148, 119-124, 1994
A:Title: Characterization and sequence analysis of a Streptomyces rochei A2 endoglucanase
A:Reference number: JC2571; MUID:95011642
A:Accession: JC2571
A:Molecule type: DNA
A:Residues: 1-382 <PER>
A:Cross-references: EMBL:X73953; NID:g393391; PIDN:CA52139.1; PID:g393392
A:Note: this cellulolytic strain was isolated from the gut of termites
C:Genetics:
A:Gene: egis
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Superfamily: bacterial cellulase-binding domain homology
C:Keywords: glycosidase; hydrolyase; polysaccharide degradation
F:1-37/Domain: signal sequence #status predicted <SID>
F:38-382/Product: endoglucanase #status predicted <MAT>
F:279-380/Domain: bacterial cellulase-binding domain homology <CB>
F:280-379/Disulfide bonds: #status predicted

Query Match 35.1%; Score 436.5; DB 2; Length 382;
Best Local Similarity 43.2%; Pred. No. 1.9e-26;
Matches 96; Conservative 30; Mismatches 89; Indels 7; Gaps 6;

QY	10	GVTSVQDEYRYVQVQENNSNOCLTINTATGAMTSTANFSGGTGAPATPSYKCH	69
DB	46	GSTVIO-GRYVQNNRMTSATQCVT--AIDSGRRVQAGOSVPTNKAPRSYSVFNGCH	102
QY	70	WGNCTTKWVGMPDIOISQASVTSMTTOYSSGAYDAVDIMTNSPTTGGQNGEIM	129
DB	103	YINC-SPTALPARISGSSAPSSISGFVDNNAVYNASDIWLDPTPTRTG-VNRTEIMI	160
QY	130	WLNSRGVQPRGSGATGATVYAGHTMNVWQOQTSKTIISVLPKATPSISNDLKAIFA	189
DB	161	WPNRGOIOPIGSQVGT-ASVAGRWVWSSGNGTNDVLSFV-APSAMSFSFVMPFVR	218
QY	190	DAARGSLNTSDYLDVNEAGFEIMOGGGLGNSFSVSVTSG	231
DB	219	ATVARGLAGMDWTLTISIQAGFEPQNGAGLAVNSFSVVTNG	260

RESULT 2
H70895

hypothetical protein RV1090 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence.revision 17-Jul-1998 #text.change 22-Oct-1999
C:Accession: H70895
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: H70895
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <COL>
A:Cross-references: GB:AL021897; GB:AL123456; NID:93256022; PIDN:CAA17206.1; PID:e125196
A:Experimental source: strain H37RV
A:Genetics:
A:Gene: RV1090

	Query Match	18.9%	Score 235.5;	DB 2;	Length 151;	
	Best Local Similarity	36.9%;	Pred. No. 2.3e-11;			
	Matches	55;	Conservative 21;	Mismatches 70;	Indels	3;
Oy	80 MPIDQISGASMTSMSTOVSSGADVAVDITNTSPPTTGGPNCGEIIMWINSRGVOP	139				
	: :					
Db	5 LPEVGQLISATSLDYNPTTGWDASTDCLDLSPTKTC-VNOCEIIMWNHOGSIOIP	63				
Oy	140 FGSQATGVAVAGHTMNWOCOOTSMKIISYLVLPFGATSINMLDKAIFADAARGLSINT	199				
	: :					
Db	64 VCSPGV-MTTIEGRKNFVVWDGSGNNMMAAYATE-PIEVMSEFDVMSFDDHATIMEPLTD	121				
Oy	200 SDYLDDVEAFETIWOGCGGGLSNFSVSZY	228				
	: :					
Db	122 SMYLRISIRAGLEPMDSGVLGVDFSSAPRY	150				

RESULT 3
 si2610
 cellulase (EC 3.2.1.4) precursor - *Aspergillus aculeatus*
 N:Alternate names: endo-1,4-beta-glucanase
 C:Species: *Aspergillus aculeatus*
 C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 21-Jul-2000
 C:Accession: SI2610; SI4118; S40186; J00458
 R:Ool, T.; Shimmyo, A.; Okada, H.; Murao, S.; Kawaguchi, T.; Arai, M.
 Nucleic Acids Res. 18, 5884, 1990
 A:Title: Complete nucleotide sequence of a gene coding for *Aspergillus aculeatus* cellulase
 A:Reference number: SI2610; MUID:9101934
 A:Accession: SI2610
 A:Molecule type: DNA
 A:Residues: 1-237 <OOI1>
 A:Cross-references: EMBL:D00546; NID:g217818; PIDN:BA00435.1; PID:g217819
 R:Ool, T.; Shimmyo, A.; Okada, H.; Hara, S.; Ikenaka, T.; Murao, S.; Arai, M.
 Curr. Genet. 10, 217-222, 1990
 A:Title: Cloning and sequence analysis of a cDNA for cellulase (FI-CMCase) from *Aspergillus*
 A:Reference number: SI4118; MUID:91064758
 A:Accession: SI4118
 A:Molecule type: mRNA
 A:Residues: 1-237 <OOI2>
 A:Cross-references: EMBL:X52525; NID:g2287; PIDN:CA36757.1; PID:g2288
 A:Accession: S40186
 A:Molecule type: protein
 A:Residues: 17-18;42-49, 'X', 51-54, 'X', 66-79;90-111,136-205, 'XX', 208-211 <OOI3>
 C:Genetics:
 A:Introns: 138/2; 212/1
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
 A:Pathway: cellulose degradation
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation; pyroglutamic acid
 F:1-16/Domain: signal sequence #status predicted <SIG>
 F:17-237/Product: cellulase #status experimental <MNT>
 F:17/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experiment

[illegible]

RESULT 4

JU0328

cellulase (EC 3.2.1.4) precursor - Erwinia carotovora subsp. carotovora

N:Alternate names: endo-1,4-beta-glucanase; endoglucanase

C:Species: Erwinia carotovora subsp. carotovora

C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 08-Oct-1999

C:Accession: JU0328

R:Sarrilahi, H.T.; Henriissat, B.; Palva, E.T.

Gene 90, 9-14, 1990

A>Title: Cells: a novel endoglucanase identified from Erwinia carotovora subsp. carotovora

A:Reference number: JU0328; MUID:90337352

A:Accession: JU0328

A:Molecule type: DNA

A:Residues: 1-264 <SAA>

A:Cross-references: GB:M32399; NID:g148389; PIDN:AAA24817.1; PID:g148390

A:Experimental source: strain SCC3193

A>Note: part of this sequence, including the amino end of the mature protein, was com

C:Genetics:

A:Gene: cels

C:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A:Pathway: cellulose degradation

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F1-32/Domann: signal sequence #status predicted <SIG>

F13-266/Product: cellulase #status experimental <MAT>

[illegible]

cellulase (EC 3.2.1.4) precursor - *Aspergillus niger*
 N:Alternate names: carboxymethylcellulase I; endo-1,4-beta-glucanase
 C:Species: *Aspergillus niger*
 C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000
 C:Accession: S55931; S60657; J50730
 R:Sakamoto, S.; Tamura, G.; Ito, K.; Ishikawa, T.; Iwano, K.; Nishiyama, N.
 Curr. Genet. 27, 435-439, 1995
 A:Title: Cloning and sequencing of cellulase cDNA from *Aspergillus kawachii* and its expr
 A:Reference number: S55931; MUID:96059347
 A:Accession: S55931
 A:Molecule type: mRNA
 A:Residues: 1-239 <SAK>
 A:Cross-references: EMBL:DJ2901; NID:g217812; PID:g217813
 A:Note: the source is designated as *Aspergillus kawachii*
 A:Accession: S60657
 A:Molecule type: protein
 A:Residues: 76-86;176-186 <SAK2>
 A:Note: the source is designated as *Aspergillus kawachii*
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
 A:Pathway: cellulose degradation
 C:Keywords: blocked amino end; glycosidase; hydrolase; polysaccharide degradation
 F:1-16/Domain: signal sequence #status predicted <Sig>
 F:17-239/Product: cellulase #status predicted <Mat>

Query Match 14.8%; Score 184.5; DB 2; Length 239;
 Best Local Similarity 29.5%; Pred. No. 3.5e-07;
 Matches 67; Conservative 27; Mismatches 84; Indels 49; Gaps 13;

OY 13 SVQDEYRVQVNEW---NSSAOCLITN-TATGA-----WTVTANFSGGTGAPATVP 62
 DB 25 SASSPYEVNOMLWCEYGTGOCYVYDKLSSGASMTTAT-----WSGEG----- 72
 OY 63 STYKCHGNCCTTK---NVGMPIC---ISOIGSANT--WSTTSSGANDVADITWNS 114
 DB 73 -----TVKSYNSGLTFDKKLVSDVSIPTSTWSDPTNQA-DVSYDLFTAA 120
 OY 115 TPTTGGPNTETIMILNSRGVOPFGSQTATGYVAHTNNVMOGGOT---SMKITSY 170
 DB 121 NMDHATSSGDYELMILWLRKYGSGVOPRIGQIAN-ATVGRKSEWVGITSTQGADEKITSF 179
 OY 171 VLTGATSTSNLDKATFADAAA-RGSINTSDYLLDVEAGEETWOGC 216
 DB 180 VAGSPINMSG-DIKDFNYTLQNGCFRPSQHLLITLQGTPTFTGG 225

RESULT 6

endoglucanase precursor [imported] - *Sulfolobus solfataricus*
 C:Species: *Sulfolobus solfataricus*
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
 C:Accession: G90360
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, V.; Allard, G.; Aweyer, M.J.; Chan
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Tml-Noc, H.P.; Redder, F
 Barrett, R.A.; Ragan, M.A.; Jensen, C.W.; Van der Oost, J.
 submitted to Genbank, April 2001
 A:Description: *Sulfolobus solfataricus* complete genome.
 A:Reference number: A99139
 A:Accession: G90360
 A:Molecule type: protein
 A:Residues: 1-334 <KUR>
 A:Cross-references: GB:AE006641; NID:g13815224; PIDN:AAK42142.1; GSPDB:GN00155
 C:Genetics:
 A:Gene: SS01949

Query Match 10.7%; Score 133; DB 2; Length 334;
 Best Local Similarity 26.7%; Pred. No. 0.0048;
 Matches 71; Conservative 39; Mismatches 100; Indels 56; Gaps 18;

OY 3 TGP---NONGVTSVQDEYRVQVNEWNS---SSAOCLITNATGANTVS---TANFSG 52

DB 81 SPPIVYINNSTVTS-----FYLEVMMNNAKTNNGNTYTFNPLTRTLVSFNLQVNPLO 135
 OY 53 GTGAPATPYSIKGCH-WGNCCTKAVNAPIQISOGSANTSW--STQVS--SGADVAY 108
 DB 136 WTNG-----YPEIYVGRKPMWTSYAGNI-FPMRIGNNTPFVWSTYILNKLDSINFIAS 190
 OY 109 DIW-----TSTPTTGTQPNQTEIMILNSRGVOPFGSQTATGV-----TVAGHTWN 156
 DB 191 DAWIVRQIAFSPETANGNDIELMWLFSG-NLOPAGQGVGEVYPIYINHTLVNATFO 249
 OY 157 VMQOQTSWKILSYV-----LTPGATIS-NLDKAI--FADAARGLN-TSDYLL 204
 DB 250 VMKKNVWGMGMEYIAFRPDGMKTYNGYVAYEPNLFKALNFA-----SYNITNYLT 303
 OY 205 DVEAGFEIMOGGGLGNSPSVSTVS 230
 DB 304 DWEFGTE-WGTWTSNGTAYFSWTISN 328

RESULT 7

adhesin AIDA-I precursor - *Escherichia coli* plasmid pIB6
 C:Species: *Escherichia coli*
 C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 08-Oct-1999
 C:Accession: S28634; S22680; S28881; S72657
 R:Benz, I.
 submitted to the EMBL Data Library, March 1992
 A:Reference number: S28634
 A:Accession: S28634
 A:Molecule type: DNA
 A:Residues: 1-1286 <BEN>
 A:Cross-references: EMBL:X65022; NID:g42254; PIDN:CAA6156.1; PID:g42255
 R:Benz, I.; Schmidt, M.A.
 Mol. Microbiol. 6, 1539-1546, 1992
 A:Title: AIDA-I, the adhesin involved in diffuse adherence of the diarrhoeagenic *Esch*
 A:Reference number: S22680; MUID:92326638
 A:Accession: S22680
 A:Status: nucleic acid sequence not shown
 A:Residues: 839-1286 <BE2>
 A:Molecule type: DNA
 A:Residues: 839-1286
 A:Cross-references: EMBL:X65022
 A:Experimental source: strain 2787
 A:Accession: S28881
 A:Molecule type: protein
 A:Residues: 50-56 <BE3>
 A:Experimental source: strain 2787
 R:Suhr, M.; Benz, I.; Schmidt, M.A.
 Mol. Microbiol. 22, 31-42, 1996
 A:Title: Processing of the AIDA-I precursor: removal of AIDA and evidence for the out
 A:Reference number: S72657; MUID:97055419
 A:Accession: S72657
 A:Molecule type: protein
 A:Residues: 847-856 <SU>
 A:Experimental source: DAEC strain 2787
 C:Genetics:
 A:Genome: plasmid pIB6
 C:Keywords: membrane protein
 F:1-49/Domain: signal sequence #status predicted <Sig>
 F:50-1286/Product: adhesin AIDA-I #status predicted <Mat>

Query Match 10.4%; Score 129; DB 2; Length 1286;
 Best Local Similarity 25.9%; Pred. No. 0.045;
 Matches 69; Conservative 32; Mismatches 93; Indels 72; Gaps 13;

OY 7 NONGVTSVQDEYRVQVNEWNS--SSAOCLITNATGANTVSANFSG----- 53
 DB 82 NSGGTQIVANNGGKTATVY-NSGSG-----NVGTGATISTIVNSGGVQRVSSGVASAT 136
 OY 54 --TGAPATY-----PSYKCHGNCCTKNNGM--PIQIS-----QISATVSMST 97
 DB 137 NLSGAGNINYNIGHASNFIIFSG---GNQITFSGIGITDSTINISGGQGVSSGVASAT 193

Db 913 SSTMSAGNGNDPFGST-----ESTLTGIVTGEISVSCSTGTITEGSTISE 960
| | | | : | | : : | | : : | | : :
QY 221 SNSFSVSATSG 231
| : | : : |
Db 961 STMTTTCVASTC 971

```

RESULT 11
T34369
hypothetical protein T19D12.1 - Caenorhabditis elegans
C1:Species: Caenorhabditis elegans
C1:date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C1:Accession: T34369

```

A; Introns: 36/1; 134/2; 180/1; 622/3; 691/2; 754/1; 1111/2; 1174/1; 1271/3; 1322/2; 1681

RESULT 12
H72240
endoglucanase - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: H72240
R:Nelson, R.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.U.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316
A:Accession: H72240
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-258 <ARN>
A:Cross-references: GB:AE001800; GB:AE000512; NID:g4982090; PIDN:AD36591.1; PID:g498205
A:Experimental source: strain MSB8
C:Genetics:
;Gene: TML524

	Query Match	8.5%;	Score 105.5;	DB 2;	Length 256;	
	Best Local Similarity	26.6%;	Pred. No. 0.48;			
	Matches	47;	Conservative	26;	Mismatches	81; Indels 23; Gaps 12.
Oy	61	YPSIYKGGH-WGNCSTTKNVGMPIQISQIG-SAVTSWSTTOVSSGADVAYADIW-TNSTP	116			
Dd	66	YPEFYGYKKPMENHNRAEGSKLPVPVSSMKSFSEVERSPDIHHBPSLPINFAMETWLTLREXY	125			
Oy	117	TTTGGPNGTEIMIMLWN-----SNGSVQ-----PESSORATGTVTAGHTNNVWGQSOTYSK	166			
Dd	126	QTEASIGDVEIMVMWEYFNNLTPGGEKIEFTIPF--VLNGSIVSG-TWELMLA-EMGND	180			
Oy	167	IISYVL-TPGATSISNLDLKAFADAAARGSNLTSDYLIDVE-AGEFLMOCGOGGS	221			
Dd	191	VLAFLKPVKPKGRKAKFDVRH-FLDPAAG-AUSSARKADEFDLFTYWEISTEERGS	235			

RESULT 13
A:82583
Conserved hypothetical protein XF2237 [Imported] - *Xyella fastidiosa* (strain 9a5c)
C:Species: *Xyella fastidiosa*
C:Date: 18-Aug-2000 #sequence,revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: A82583
R:Anonymous, The *Xyella fastidiosa* Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen *Xyella fastidiosa*.
A:Reference number: A82515, MUID:20365717
A>Note: For a complete list of authors see reference number A59328 below
A:Accession: A82593
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-958 <SIM>
A:Cross-references: GB:AE004036; GB:AE003849; NID:9107384; PIDD:AAF85036.1; GSPDB:GN93
A:Experimental source: strain 9a5c
R:Stimpson, A.J.G.; Reisch, R.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
B:Ribeiro, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre
as-Melo, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to Genbank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurame, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
A:Genetics:
A:Gene: XF2237

```

Query Match          8.5%; Score 105.5; DB 2; Length 958;
Best Local Similarity 20.4%; Pred. No. 2.1;
Matches 56; Conservative 38; Mismatches 93; Indels 87; Gaps 11;

QY      13 SVQDEYNQTNEMNSSAQOCLTTITATGAMTVSTANFSGTGAPATPSYIK-----66
      111:::::::::1:
Db      606 SFADCKYLLR-GGNN-----RAIRAPISITELYEPNTIALMDGSDPCAGTSPFTQACANT 660
      :::::

QY      67 ---GCHMGNCCTKNKMPLOIS-----OLGSAVTSMTSTQVSSGADVAV- 108
      :|||:::111
Db      661 GLAAQYGVNTANPFGQYNQISGQNQLKPEIADTWITLGFALTPIPKDDLDLSADYISIMKK 720
      :|||:::111

QY      109 -----DIWTSPTPTTKGQPNGEIMIMLSR 134
      ::::1:111:1111

Db      721 DTIREIGASNILLTACALINDANILCSRIIRRNSTGTGDLFLGNSPATSG-----LVLSNR 772
      ::::1:111:1111

QY      135 GGVQVFGSQTATGTVY-AGHTWNNWQGOOTSKIISYVL-----TPGATISNLDLKA1 187
      111:::1:111:1111:1111

Db      773 GN---FGALQRFEGIDILTYSIANNVWPGRLTITMGLNIYIKODYQVPG-ISENNISCAGI 828
      :111:::1111:1111:1111

QY      188 FADAAA-----RGLNNTSDYLLDIVEAGEFEIWWCG 216
      :111:::1111:1111:1111

```

Db 829 VNTACGADGRRKGSGSKRMHVMVNRGDFRTVG 862

RESULT 14

A29832

HPI layer surface protein precursor - Deinococcus radiodurans

C:Species: Deinococcus radiodurans

C>Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 15-Oct-1999

C:Accession: A29832

R:Peterson, J.; Peterson, M.; Lottspeich, F.; Schaefer, W.; Baumeister, W.

J. Bacteriol. 169, 5216-5223, 1987

A:Title: Nucleotide sequence analysis of the gene encoding the Deinococcus radiodurans

A:Reference number: A29832; MUID:86032846

A:Accession: A29832

A:Molecule type: DNA

A:Residues: 1-1036 <PEPT>

A:Cross-references: GB:M17895; NID:G145035; PIDN:AAA23335.1; PID:G145036

C:Comment: Although about 0.7 mol of 16:1 fatty acid was recovered per mol of blocked am

C:Genetics:

A:Gene: hpi

C:Keywords: blocked amino end; lipoprotein

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19/Binding site: sn-2,3-diacylglycerol (Cys) (covalent) #status predicted

F:19/Modified site: fatty acylated amino end (Cys) (in mature form) #status predicted

F:74-86,642-754/Disulfide bonds: #status predicted

F:256-275/Disulfide bonds: #status predicted

Query Match 8.4%; Score 105; DB 2; Length 1036;

Best Local Similarity 24.9%; Pred. No. 2.5; Mismatches 111; Indels 44; Gaps 11;

Matches 62; Conservative 32; Mismatches 111; Indels 44; Gaps 11;

Db 10 GVT-----VQGEYVQVEMNSAQCCTTNTATGAMVTSTANFSGGTG----- 55

144 GVSAPITFKDANGNVQGYDNYVDNATITVARGVYTVAGNVSGFPTTNRDLS 203

56 GAPATYPSIYKCGHNGCTKNVG-MPIQISQISAVTSMSTQVSSGAYDAVD--WT 112

204 GGNQVTVTLNTQA--GTTTPRPVGSINITPAVGTSTVIGSTVRV--FDKANEVQCMV 258

113 NSTPTTGQPNCTE--IMIMLSRKGVOPEFSQTATGTVAGHTVMWMOGQOTSMKITS 169

259 GGAAAVTAQVDSTSGCDVVPVNSGTNV---VITVMKGVGNGQVYATARNISVQAAS 314

170 Y-VLRPGATISNLDKALFADA-----AARGSLNTSDYLLDEAGFEIMOGGGLGSN 222

315 YGVVTPAGDQELTLTSEGIVRADSGWMRLGQGVSTPSPDNLNLDIYK-----GTV 366

223 SPSVSTSG 231

367 NFSVNPAG 375

RESULT 15

BA7093

cellulase (EC 3.2.1.4) Cend - Cellulomonas fimi

N:Alternate names: endo-1,4-beta-glucanase

C:Species: Cellulomonas fimi

C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 06-Dec-1996

C:Accession: BA7093

R:Meinke, A.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.

J. Bacteriol. 175, 1910-1918, 1993

A:Title: Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase D (Cend),

A:Reference number: A47093; MUID:93209933

A:Accession: BA7093

A:Status: preliminary

A:Molecule type: DNA; protein

A:Residues: 1-747 <MEIT>

A:Experimental source: ATCC 484

A>Note: sequence inconsistent with the nucleotide translation

A>Note: sequence extracted from NCBI backbone (NCBIN:128120, NCBI:128122)

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A:Pathway: cellulose degradation

C:Superfamily: bacterial cellulose-binding domain homology

C:Keywords: glycosidase; hydrolyase; polysaccharide degradation

F:644-745/Domain: bacterial cellulose-binding domain homology <BCB>

F:645-744/Disulfide bonds: #status predicted

Query Match 8.4%; Score 104; DB 2; Length 747;

Best Local Similarity 23.9%; Pred. No. 2.1;

Matches 60; Conservative 30; Mismatches 91; Indels 70; Gaps 11;

Db 20 RQTEEMNSAQCCTTNTATGAMVTSTANFSGGTGAPATYPSIYKCGHNGCTKNVG 79

521 RAKDAGNTSA-----ASAAVYARJA----AGGDVTPASVPTGLTACTPTATSV 567

80 MPIQISQISAVTSMSTQVSSGAYDAVDIWTNSTPTTGQPNCTEIMW----- 130

568 L-----TWASTDTGSGSVTGYEYRGS--ILVAPRTGSHITGSLAATAYT 613

131 -----LNSRQVQ-----PGSQTA----TGYTVAGHTVMWMOGQOTSMKITSYVLT-PGA 176

614 FTVRAVDAGNVSAASAPVGTTPADPTTGSQAVTYTANGMGCFPA----AVTLFNTGT 669

177 TSISNLDKALF-----ADAAARGSLNTS-----DYLLVEAGFEIMOGGGLG 220

670 TALSGWTLGFAFPSSQOTLTQGSARMAGSSSVTATNEAMNAVLAAGASVEIGFGTHTG 729

221 SNSFSVSTSG 231

730 TMTATPTFTVG 740

Search completed: August 29, 2002, 16:24:16

Job time: 351 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:38:45 ; Search time 38.44 Seconds

(Without alignments)
232.680 Million cell updates/sec

Title: US-09-917-384-7

Perfect score: 1244
Sequence: 1 DCTPGPNNGVTYSVODGEYR.....IMCGGGLGINSFSVSTSG 231

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	234	18.8	237	1 GUN_ASPAC
2	215	17.3	264	1 GUN_ERMCA
3	184.5	14.8	239	1 GUN_ASPAC
4	129	10.4	1286	1 AIDA_ECOLI
5	112	9.0	1025	1 PMP_CILMU
6	105	8.4	1036	1 HP12_DEIRA
7	104	8.4	747	1 GUN_CELFI
8	98	7.9	3075	1 LMAI_HUMAN
9	96.5	7.8	948	1 HP11_DEIRA
10	96.5	7.8	1076	1 NUP1_YEAST
11	96	7.7	479	1 CBSA_SULSO
12	96	7.7	1475	1 N153_HUMAN
13	95	7.6	211	1 YOR3_SOUV3
14	95	7.6	666	1 MUR2_ENTHR
15	94.5	7.6	1251	1 Y003_CAEEL
16	94	7.6	657	1 CSPI_CORGL
17	93.5	7.5	380	1 PGL1_SCLSC
18	93	7.5	335	1 XYNB_SRLI
19	92.5	7.4	863	1 YEJO_ECOLI
20	92	7.4	313	1 KREL_YEAST
21	91	7.3	197	1 XYNB_SCHCO
22	91	7.3	525	1 NUG2_RAT
23	91	7.3	550	1 MYCO_STRIC
24	90.5	7.3	2124	1 PGCA_CAP
25	89.5	7.2	796	1 YSBA_CAEEL
26	89	7.2	610	1 CHIT_STRPL
27	89	7.2	718	1 CDGT_BACCI
28	88.5	7.1	697	1 CRAC_DICDI
29	88	7.1	225	1 XYN2_ASPNG
30	88	7.1	430	1 NUA2_YEAST
31	88	7.1	598	1 YAU6_SCHPO
32	88	7.1	718	1 CDGT_BACSS
33	87.5	7.0	159	1 MP63_MYCTU

34	87.5	7.0	530	1 YAG_A_SCHPO	Q09768 schizosacch
35	87.5	7.0	773	1 CDH_PHACH	Q01738 phanerochae
36	87.5	7.0	887	1 QUA1_DROME	Q23969 drosophila
37	87	7.0	225	1 XYN1_EMENT	P55332 emericella
38	87	7.0	605	1 YHC8_YEAST	P38739 saccharomyc
39	87	7.0	1447	1 VGL2_CVPT	Q01977 porcine tra
40	87	7.0	1449	1 VGL2_CVPT	P18450 porcine tra
41	86.5	7.0	713	1 CDG2_PAEBA	P31835 penicillium
42	86.5	7.0	966	1 FIB1_PETMA	P02674 petromyzon
43	86.5	7.0	1025	1 SIAP_CAUCR	P35828 caulibacter
44	86	6.9	886	1 VGP3_EBYA8	Q07264 epstein-bar
45	86	6.9	1045	1 GUNB_CELFI	P26225 cellulomona

ALIGNMENTS

```

RESULT 1
GUN_ASPAC 1
ID GUN_ASPAC STANDARD; PRT: 237 AA.
AC P22669;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Endoglucanase I precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase) (FI-CMCase).
OS Aspergillus aculeatus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5053;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F-50;
RX MEDLINE=91016934; PubMed=2216782;
RA Ooi T., Shimmyo A., Okada H., Murao S., Kawaguchi T., Arai M.;
RT "Complete nucleotide sequence of a gene coding for Aspergillus
RL aculeatus cellulase (FI-CMCase).";
RL Nucleic Acids Res. 18:5884-5884(1990).
RN [2]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC STRAIN=F-50;
RX MEDLINE=91064758; PubMed=2249253;
RA Ooi T., Shimmyo A., Okada H., Hara S., Ikenaka T., Murao S.,
RA Arai M.;
RT "Cloning and sequence analysis of a cDNA for cellulase (FI-CMCase)
RL from Aspergillus aculeatus";
RL Curr. Genet. 18:217-222(1990).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- INDUCTION: BY CELLULOSIC MATERIALS AND HEMICELLULOSES.
CC -1- MISCELLANEOUS: WILL ALSO HYDROLYSE 1,4-LINKAGES IN BETA-D-GLUCANS
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY H (FAMILY 12 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL, D00546; BA00435.1; -;
DR EMBL, X52525; CA36757.1; -;
DR PIR, JQ0458; JQ0458.
DR PIR, J2610; J2610.
DR PIR, S1418; S1418.
DR InterPro: IPR002594; Glyco_hydro_12.
DR Pfam: PF01670; Glyco_hydro_12; 1.
DR ProDom: PD004316; Glyco_hydro_12; 1.
KW Cellulose degradation; Hydrolyase; Glycosidase; Signal.

```

FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 237 ENDOGLUCANASE 1.
FT MOD_RES 17 17 PYRROLIDONE CARBOXYLIC ACID.
SQ SEQUENCE 237 AA: 25560 MW: 8f173571a8a6931 CRC64;

Query Match 18.8%; Score 234; DB 1; Length 237;
Best Local Similarity 34.7%; Pred. No. 8.3e-12;
Matches 76; Conservative 24; Mismatches 99; Indels 20; Gaps 10;

OY 17 GERYVQTNEMN---SSAOCLTINTATGAMTV-STA-NFSGTGAPATYPSIYKCHWG 71
DB 31 GYITINNIMMGKADAGSGSCITVNSASAGTSMSTKMNNGSGENSVK-----YAN 81
OY 72 NOTTNVNGPIQISQISAVTSMSTTOVSSGAYVADYWTNSTPTTGGPNGTEIMWL 131
DB 82 SGLTFNKKLVLSQISQI-PTTARMSYDNTGIRA-DVAAYDFTADINHTWAGSDYELIMWL 139
OY 132 NSRGVQVPGSQATGATVAGHTWVWVWOGQTSWKIISVYLTPGATISNLDLKAIFADA 191
DB 140 ARYGVQPIGSIQIAT-ATYDQGTWELWTCANGSOKTYSFAVAPTPITTSFOG-DVNDFFKYL 197
OY 192 AA-RGSLNTSDYLDVGEAFELWOGQG-LGSNSFSVSV 228
DB 198 TQNHGPASSQYLITLQFCTEPTGPGPATLSVNSASV 236

RESULT 2

GUNS_ERMCA STANDARD; PRT: 264 AA.
AC P16630:

DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase S precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase S)
GN CELS.

OS Erwina carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=554;

RM SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN-SCG193;

RA MEDLINE-90337352; PubMed-2379837;

RA Saarihanen H.T., Hemmisaari B., Palva E.T.;

RT "Cels: a novel endoglucanase identified from Erwina carotovora
subsp. carotovora.";

RL Gene 90:9-14(1990).

CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY H (FAMILY 12 OF GLYCOSYL
HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

CC EMBL: M32399; AAA24817.1; -.

DR PIR: J00328; J00328.

DR InterPro: IPR002594; Glyco_hydro.12.

DR Pfam: PF01670; Glyco_hydro.12; 1.

DR ProDom: PD004316; Glyco_hydro.12; 1.

KW Cellulose degradation; Hydrolase; Glycosidase; Signal.

FT SIGNAL 1 32

FT CHAIN 33 264

SQ SEQUENCE 264 AA: 29757 MW: E6D61388950C77AA CRC64;

Query Match 17.3%; Score 215; DB 1; Length 264;
Best Local Similarity 30.5%; Pred. No. 3e-10;
Matches 67; Conservative 33; Mismatches 90; Indels 30; Gaps 11;

OY 16 DGEYVQTNEMNNSA---QOCLTINTATGA-----WTVSTANFSGTGAPATYPSIYK 66
DB 46 NKKYVLFNNWKGKDELKMGQGITIFNPSISMGNMNMPSSTHSVK-----ATYSLVS 97
OY 67 GCHWNGCTTKVNGMPQISQISAVTSMSTTOVSSGAYVADYWTNSTPTTGGPNGT- 125
DB 98 GHWHTAGTYENSGLPQLSSNKSITSNVTYSIKATGTYNAADVIMFHTDKANMDSPPND 157
OY 126 EIMIMLNHSGVQVPGSQATGATVAGHTWVWVWOG-----QGTSMKTIISVYLTPGATIS 180
DB 158 ELMTIMLNDT-NAGPADYLET-VFLGDSMNVPFKGINADNGGWNVFSVHTSGTNSAS 215
OY 181 NDLKAIFADAAARGSLNTSD--YLLDVEAGEFELWOGQG 218
DB 216 -LNIRH-FTDYLVTQKMSDEKIVISVEFGTEIF-GGSG 252

RESULT 3

GUNA_ASPAK STANDARD; PRT: 239 AA.
AC Q12679:

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)
GN CERK.

OS Aspergillus awamori (var. kawachi).

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

OX NCBI_TaxID=105351;

RM SEQUENCE FROM N.A.

RC STRAIN-IFO 4308;

RA MEDLINE-96059347; PubMed-7586029;

RA Sakamoto S., Tamura G., Ito K., Ishikawa T., Iwano K., Nishiyama N.;

RT "Cloning and sequencing of cellulase cDNA from Aspergillus kawachi
and its expression in Saccharomyces cerevisiae.";

RL Curr. Genet. 27:435-439(1995).

CC -1- FUNCTION: HAS CARBOXYMETHYLCELLULOSE ACTIVITY.

CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY H (FAMILY 12 OF GLYCOSYL
HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

CC EMBL: D12901; BAA02297.1; -.

DR InterPro: IPR002594; Glyco_hydro.12.

DR Pfam: PF01670; Glyco_hydro.12; 1.

DR ProDom: PD004316; Glyco_hydro.12; 1.

KW Cellulose degradation; Hydrolase; Glycosidase; Signal.

FT SIGNAL 1 16

FT CHAIN 17 239

SQ SEQUENCE 239 AA: 25769 MW: 16B0304A138276D0 CRC64;

Query Match 14.8%; Score 184.5; DB 1; Length 239;
Best Local Similarity 29.5%; Pred. No. 7e-08;
Matches 67; Conservative 27; Mismatches 84; Indels 49; Gaps 13;

OY 13 SYVDGGEYVQTNEMN---NSAOCLTINTATGA-----WTVSTANFSGTGAPATYTP 62
DB 13 SYVDGGEYVQTNEMN---NSAOCLTINTATGA-----WTVSTANFSGTGAPATYTP 62


```

Db 25 SASSPPSYNONIMGEYQGTGSCVTVYDKLSSSGASWHTKW-----WSGEG----- 72
QY 63 STYKGCCHMCNCTTK-----NYGMPIQ-----ISOIGSAVTS--WSTTOVSSGAVDAVDTWNS 114
Db 73 -----TVKSYNSNGSLFPDKRLVDVSIPTSVWMSQDDTQVQA-DVSYDLFTAA 120
QY 115 PTTTGGPQNGTEIMTWLNSRGVQVPGSQTPATGTVAGHTWVWQOQT-----SMKIISY 170
Db 121 NADHATSSGSDYELMIAWLGSGVQPIGKQIAT-ATVGGKSMEWYGTSTQAGAEQKTYSF 179
QY 171 VLTPTGATSTSNIDKAIKAFDAAA-RGSLMTSDVLLDVEAGFEIWMOG 216
Db 180 VAGSPINWSWG-DIKDFNYLTQNGFPASSOHLITLQCTEPTTGG 225

RESULT 4
AIDA_ECOLI STANDARD; PRT: 1286 AA.
ID AIDA_ECOLI
AC 003135;
DT 01-JUN-1994 (Rel. 29, Created)
DR 01-JUN-1994 (Rel. 29, Last sequence update)
DE 01-JUN-1994 (Rel. 29, Last annotation update)
DE Adhesin aida-1 precursor.
GN AIDA-I.
OS Escherichia coli.
OG Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 50-56.
RX STRAIN-2787 (O126:H27);
RX MEDLINE-92326638; PubMed-1625582;
RA Benz I., Schmidt M.A.;
RT "AIDA-I, the adhesin involved in diffuse adherence of the
RT diarrhoeagenic Escherichia coli strain 2787 (O126:H27), is
RT synthesized via a precursor molecule."
RL Mol. Microbiol. 6:1539-1546(1992).
CC -1- FUNCTION: THIS IS AN ADHESION PROTEIN NECESSARY FOR THE DIFFUSE
CC ADHERENCE OF DIARRHEA-CAUSING ENTEROPATHOGENIC ESCHERICHIA COLI
CC TO EPITHELIAL CELLS.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X65032; CAA46156.1; -.
DR PIR: S28634; S28634.
KW Cell adhesion; Signal; Outer membrane; Plasmid.
FT SIGNAL 1 49
FT CHAIN 50 1286 ADHESIN AIDA-I.
FT PROPEP 1286 1286
SO SEQUENCE 1286 AA; 132271 MW; B2A00F72AC05FB34 CRC64;

Query Match 10.4%; Score 129; DB 1; Length 1286;
Best Local Similarity 25.9%; Pred. No. 0.011;
Matches 69; Conservative 32; Mismatches 93; Indels 72; Gaps 13;

QY 7 NONGCVTSYODGERVQOTNEMNSSAOCLTINTATGATVSTANFSG----- 53
Db 82 NSGGTQIVNNGGKTTATTV-NSSGSO-----NNGTSATISTIVNSGCIORVSSGVAASAT 136
QY 54 --TGCAPAY-----PSIYKCHMGNCCTTKNVGM--PIQIS-----OIGSAVTSWSTP 97
Db 137 NLSGGMNINYNLGHASNYVIFSG--GNQITFGCGITDSTINISSGGQQRKSSGCVASNTT 193
QY 98 QVSSGAVDAVDAVDTWNSPTTTC-----QPNGTEIMTWLNSRGVQVPGSQ--QTATGTV 149

```

```

Db 194 INSSGAMNIISEGAISTHISGNGNOYISAGANATERIV---NSGGFORVNSGAVATGTV 250
QY 150 VAGHTWVWQOQTQSMKIISYVLTPTGATSTSNIDKAIKAFDAA-----ARGLN----- 198
Db 251 LSGGTQVNSGSGSAI-----STVYNSGVQTVFAGATVDTFTVNSGGMNIISSG 299
QY 199 --TSVDLLDVEAGFEIWMOGGQGLSN 222
Db 300 GIVSEFTVANSQTONIYSSGSAISAN 325

RESULT 5
PMPF_CHLMU STANDARD; PRT: 1025 AA.
ID PMPF_CHLMU
AC 09P146;
DT 16-OCT-2001 (Rel. 40, Created)
DR 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable outer membrane protein pmpF precursor (Polymorphic membrane
DE protein F).
GN PMPF OR TCO262.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-MOPn / N199;
RX MEDLINE-20150255; PubMed-10684935;
RX Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RX White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
RX Linher K., Weldman J., Khouri H., Craven B., Bowman C., Dodson R.,
RX Gysin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RX Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MOPn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE002293; AAF39131.1; -.
DR TIGR: TC0262;
DR InterPro: IPR003357; OMP.
DR Pfam: PF02385; OMP; 1.
KW Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 1025 PROBABLE OUTER MEMBRANE PROTEIN PMPF.
SO SEQUENCE 1025 AA; 111458 MW; EA472E07818B9PF2 CRC64;

Query Match 9.0%; Score 112; DB 1; Length 1025;
Best Local Similarity 22.6%; Pred. No. 0.19;
Matches 61; Conservative 25; Mismatches 94; Indels 90; Gaps 12;

QY 7 NONGCVTSYODGERVQOTNEMNSSAOCLTINTATGATVSTAN-----FSGGTGAPA 59
Db 536 NDOGTTAP-----RIRN-----PDMTQNTNSNQAPVSTENVAQKIFSGVLSLVDE 564
QY 60 TYSIYKGCCHMGNCCTTKNVGMPIQISOIGSAVTSWSTTOVSSGAVDAV---DIPTNST 115
Db 585 NYESVYDSCOLSGKANKOPLIETITNDALSDMKRT-LNTSLYSLPHRYGGLWTSNW 643
QY 116 PTTT---GQPNTEIMTWL-----SRGGVQVPGSQOTA 145

```



```

FT DOMAIN 1150 1159 LAMININ EGF-LIKE 14 (N-TERMINAL).
FT DOMAIN 1160 1361 LAMININ DOMAIN IV 2 (DOMAIN IV A).
FT DOMAIN 1162 1402 LAMININ EGF-LIKE 14 (C-TERMINAL).
FT DOMAIN 1403 1451 LAMININ EGF-LIKE 15.
FT DOMAIN 1452 1508 LAMININ EGF-LIKE 16.
FT DOMAIN 1509 1555 LAMININ EGF-LIKE 17.
FT DOMAIN 1556 2116 DOMAIN II AND I.
FT DOMAIN 2117 2297 LAMININ G-LIKE 1.
FT DOMAIN 2305 2481 LAMININ G-LIKE 2.
FT DOMAIN 2486 2673 LAMININ G-LIKE 3.
FT DOMAIN 2713 2885 LAMININ G-LIKE 4.
FT DOMAIN 3070 2890 LAMININ G-LIKE 5.
FT DOMAIN 1706 1795 COILED COIL (POTENTIAL).
FT DOMAIN 1968 2088 COILED COIL (POTENTIAL).
FT DOMAIN 2120 2534 COILED COIL (POTENTIAL).
FT SITE 2534 2536 CELL ATTACHMENT SITE.
FT DISULFID 270 279 BY SIMILARITY.
FT DISULFID 272 290 BY SIMILARITY.
FT DISULFID 292 301 BY SIMILARITY.
FT DISULFID 297 305 POTENTIAL.
FT DISULFID 304 324 BY SIMILARITY.
FT DISULFID 327 336 BY SIMILARITY.
FT DISULFID 329 361 BY SIMILARITY.
FT DISULFID 364 373 BY SIMILARITY.
FT DISULFID 376 394 BY SIMILARITY.
FT DISULFID 397 409 BY SIMILARITY.
FT DISULFID 399 427 BY SIMILARITY.
FT DISULFID 429 438 BY SIMILARITY.
FT DISULFID 441 451 BY SIMILARITY.
FT DISULFID 454 467 BY SIMILARITY.
FT DISULFID 466 471 BY SIMILARITY.
FT DISULFID 473 482 BY SIMILARITY.
FT DISULFID 485 500 BY SIMILARITY.
FT DISULFID 742 751 BY SIMILARITY.
FT DISULFID 744 757 BY SIMILARITY.
FT DISULFID 760 769 BY SIMILARITY.
FT DISULFID 772 788 BY SIMILARITY.
FT DISULFID 791 806 BY SIMILARITY.
FT DISULFID 793 816 BY SIMILARITY.
FT DISULFID 819 828 BY SIMILARITY.
FT DISULFID 831 846 BY SIMILARITY.
FT DISULFID 849 863 BY SIMILARITY.
FT DISULFID 851 870 BY SIMILARITY.
FT DISULFID 873 882 BY SIMILARITY.
FT DISULFID 885 899 BY SIMILARITY.
FT DISULFID 902 914 BY SIMILARITY.
FT DISULFID 904 921 BY SIMILARITY.
FT DISULFID 923 932 BY SIMILARITY.
FT DISULFID 935 948 BY SIMILARITY.
FT DISULFID 951 963 BY SIMILARITY.
FT DISULFID 953 969 BY SIMILARITY.
FT DISULFID 971 980 BY SIMILARITY.
FT DISULFID 983 995 BY SIMILARITY.
FT DISULFID 998 1007 BY SIMILARITY.
FT DISULFID 1000 1014 BY SIMILARITY.
FT DISULFID 1016 1025 BY SIMILARITY.
FT DISULFID 1028 1041 BY SIMILARITY.
FT DISULFID 1044 1056 BY SIMILARITY.
FT DISULFID 1065 1074 BY SIMILARITY.
FT DISULFID 1074 1087 BY SIMILARITY.
FT DISULFID 1077 1087 BY SIMILARITY.
FT DISULFID 1403 1412 BY SIMILARITY.
FT DISULFID 1405 1419 BY SIMILARITY.
FT DISULFID 1422 1431 BY SIMILARITY.
FT DISULFID 1434 1449 BY SIMILARITY.
FT DISULFID 1452 1466 BY SIMILARITY.
FT DISULFID 1454 1476 BY SIMILARITY.
FT DISULFID 1479 1488 BY SIMILARITY.
FT DISULFID 1491 1506 BY SIMILARITY.
FT DISULFID 1509 1521 BY SIMILARITY.
FT DISULFID 1511 1528 BY SIMILARITY.
FT DISULFID 1530 1539 BY SIMILARITY.
FT DISULFID 1542 1553 BY SIMILARITY.

```

```

FT DISULFID 1556 1556 INTERCHAIN (PROBABLE).
FT DISULFID 1560 1560 INTERCHAIN (PROBABLE).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 555 555 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 665 665 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 763 763 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 952 952 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1045 1045 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1407 1407 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1579 1579 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1596 1596 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1678 1678 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1689 1689 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1698 1698 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1717 1717 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1804 1804 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1894 1894 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 7.9%; Score 98; DB 1; Length 3075;
Best Local Similarity 22.1%; Pred. No. 7.9;
Matches 66; Conservative 40; Mismatches 108; Indels 84; Gaps 15;

QY 9 NGVTSVQGEYVQCNEMNS-----SAQCCLTINTATGATV----- 45
DB 2184 SSTRLEPPDPFIDNRWHSIHVAPNGISLSYKEMSNQSPKTSKSPETAVLDVN 2243
QY 46 -STANFSGGTG---APATPSIYKCHWNC---TTKNVAPIQISQISAVTSMTQ 98
DB 2244 NSTLNFVGGIGQIKKSPAVKVTHERKGC-LGEAFNGISIGIMNIERKGRGCGFSQ 2302
QY 99 VSSGAYDV---AYDIWNTSPPTTQO-----PNTGEIMWINS-----R 134
DB 2303 NEDPSFHEPDSGYVEKSLPATVTOIIMLFNTFSPNG--LLLYGYSYKDFLSTIEFR 2360
QY 135 GGVO---PPGSGFATGVYAGHTMWV-----QGTSKIIISVILP 174
DB 2361 GRVYKWTDLGSGPITLLDRRNNGTWTYKIAFORNRKQGLAVDAVYNSNETROGETP 2420
QY 175 GATG-TISNLDKAIPIA-----DMAARGSLNTSDYLLDEAGFEIMOGGGLGSFSFV 226
DB 2421 GASDLDNRDKRPIYVGGIPRSRVRRGVTTKSPFGCIK-NLEISRSFDDLNRNGV 2477

RESULT 9
HP1L DEIRA STANDARD; PRT; 948 AA.
AC P56867;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hexagonally packed intermediate-layer surface protein precursor.
GN HPI OR DR2508.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RL;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Ueberlack T., Zalewski C.,
RA Makareva K.S., Atavind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.,
RA "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RL."
RL Science 286:1571-1577(1999).
CC -!- FUNCTION: SHAPE MAINTENANCE, POSSIBLE PROTECTION FROM NOXIOUS
CC ENZYMES OR EXOGENOUS AND UNSETTLING DNA, AND MAY MEDIATE HOMOTYPIC
CC CELL-CELL CONTACTS (BY SIMILARITY).

```

```

OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=90275616; PubMed=2190694;
RA Davis L.I., Fink G.R.;
RT "The NUP1 gene encodes an essential component of the yeast nuclear
   pore complex.";
RL Cell 61:965-978(1990).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97344368; PubMed=9200815;
RA Voss H., Benes V., Andrade M.A., Valencia A., Rechmann S., Teodoru C.,
RA Schwager C., Paces V., Sander C., Ansorge W.;
RT "DNA sequencing and analysis of 130 kb from yeast chromosome XV.";
RL Yeast 13:655-672(1997).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSPORT.
CC THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.
CC -1- SUBCELLULAR LOCATION: Nuclear pore complex.
CC -1- DOMAIN: APPEARS TO BE DIVIDED INTO THREE DOMAINS DEFINED BY
CC CENTRALLY LOCATED REPEATING UNITS. FUNCTIONAL N-TERMINAL OF NUP1
CC OR OF NUP2 IS REQUIRED FOR GROWTH.
CC -1- DOMAIN: CONTAINS X-F-X-F-G REPEATS.
CC -1- SIMILARITY: TO THE CENTRAL REPEATING UNITS OF NUP1 AND NUP2, AND
CC TO MAMMALIAN P62.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M33632; AAA34822.1; -
DR EMBL; X94335; CAA64020.1; -
DR EMBL; Z75006; CAA99295.1; -
DR PIR; A35622; A35622.
DR SGD; S0005624; NUP1.
RW Nuclear protein; Transport; Repeat.
KW DOMAIN 333 949 29 x 9 AA APPROXIMATE REPEATS.
FT DOMAIN
SQ SEQUENCE 1076 AA; 113581 MW; 4AC23567D2FB53CC CRC64;
-----
Query Match 7.8% Score 96.5; DB 1; Length 1076;
Best Local Similarity 23.9%; Pred. No. 3.3;
Matches 47; Conservative 27; Mismatches 70; Indels 53; Gaps 8;
QY 27 NSSAOCCLTITATGAMVSTANFSGGTGAGAPATVPYSLYKCHGNGCTTK----- 76
Db 793 NNTSSAFSPFCTAMTNGTANASANSISFENPAT----- GNGTTTSNMSGTINIA 842
QY 77 ---AVGAMIQISQIGSAVTSMTSTOVSSGADVADYDITNSTPTTGTGQPKCTEIMIMNS 133
Db 843 GTFVNGKDDQ-----STASGNTGAGS-----AFGSSSGTAATGGAASNOSSEFNGNG 891
QY 134 RGVQVQFSSQATGATGVTVG-----HTVMWQGGQSTSMKIIISVLTGATISLNT-- 182
Db 892 AGGLNPFISATSSSTNANAGLFNKPSTNAQNV--NVPSAFNFTGNNSTPGGSGTFMANGN 949
QY 183 -DLKAIIFADAAAGSLN 198
Db 950 TNANTVEA-----GSNN 961
-----
RESULT 11
ID CBSA_SULSO STANDARD; PRT; 479 AA.
AC P58029;
DT 16-OCT-2001 (Rel. 40, Created)

```

DR 16-OCT-2001 (Rel. 40, last sequence update)
 DT 01-MAR-2002 (Rel. 41, last annotation update)
 DE Cytochrome b558/566 subunit A.
 GN CESA OR SSO2801 OR C48_011.
 OS Sulfolobus solfataricus.
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 OX NCBI_TaxID=2287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=21332296; PubMed=11427726;
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Meyer M.J., Chan-Welher C.C.-Y., Clausen I.G., Curtis B.A.,
 RA De Moers A., Erasmo G., Fletcher C., Gordon P.M.K.,
 RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 CC -1- FUNCTION: MONOHEME CYTOCHROME WHOSE PHYSIOLOGICAL FUNCTION IS NOT
 CC -1- YET CLEAR (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AE006874; AAK2914.1.
 KW Electron transport; Transmembrane; Heme; Glycoprotein;
 KW Complete proteome.
 FT TRANSMEM 19 39 POTENTIAL.
 FT CARBOHYD 456 476 POTENTIAL.
 FT CARBOHYD 73 73 N-LINKED (POTENTIAL).
 FT CARBOHYD 99 99 N-LINKED (POTENTIAL).
 FT CARBOHYD 152 152 N-LINKED (POTENTIAL).
 FT CARBOHYD 172 172 N-LINKED (POTENTIAL).
 FT CARBOHYD 182 182 N-LINKED (POTENTIAL).
 FT CARBOHYD 191 191 N-LINKED (POTENTIAL).
 FT CARBOHYD 219 219 N-LINKED (POTENTIAL).
 FT CARBOHYD 262 262 N-LINKED (POTENTIAL).
 FT CARBOHYD 288 288 N-LINKED (POTENTIAL).
 FT CARBOHYD 302 302 N-LINKED (POTENTIAL).
 FT CARBOHYD 325 325 N-LINKED (POTENTIAL).
 FT CARBOHYD 348 348 N-LINKED (POTENTIAL).
 FT CARBOHYD 385 385 N-LINKED (POTENTIAL).
 SQ SEQUENCE 479 AA; 52532 MW; 7C92B0B91A8245B4 CRC64;
 Query Match 7.7%; Score 96; DB 1; Length 479;
 Best Local Similarity 24.9%; Pred. No. 1.5; Mismatches 79; Indels 76; Gaps 14;
 Matches 58; Conservative 20;
 OY 4 PEPNONGVTSDGDERVO---TNEWSSAOCLTINTAGTAVTSTANFSGGTGAGAPA 59
 DB 252 PQAFDFGVDFTDGGSLVPGGANIM-----MW-VSGATWNN-----A 289
 OY 60 TYPSTIYKGGHMGNCYTKNKGMP-IQISQISGAVTSMSTVOVSSGAYD-AYDIWINSPTT 117
 DB 290 TYDPAKFWLWON--TSLTGLPYIDPDNHGFAVPLYTN--WTNNYEVDYTAGIM--YTPV 342
 OY 118 TGPONGTFTIMILNRSRGVQPFQSGTATGTVAGTVMWMOGQOTSMKTIISYVLTGAT 177
 DB 343 TTSGLNGSLFTW-----TGATYQNGYTWV-----EFARPLAVPSA- 378
 OY 178 SISNDLKAIFAADAAAGSLNTSYDLLDVEAGFEIMWOGGQGLGNSFSVSYS 230
 DB 379 -----YAKMMPNITVGTYYV-----AFAYWOG--KLGFTLFDKSTIS 414

RESULT 12
 ID N153_HUMAN STANDARD: PRT; 1475 AA.
 AC P49790;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Nuclear pore complex protein Nup153 (Nucleoporin Nup153) (153 kDa
 DE nucleoporin).
 GN NUP153.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=94154002; PubMed=8110839;
 RA McMorris I., Bastos R., Horton H., Burke B.;
 RT "Sequence analysis of a cDNA encoding a human nuclear pore complex
 RT protein, hnup153.";
 RL Biochim. Biophys. Acta 1217:219-223(1994).
 CC -1- FUNCTION: POSSIBLE DNA-BINDING SUBUNIT OF THE NUCLEAR PORE
 CC COMPLEX (NPC). THE REPEAT-CONTAINING DOMAIN MAY BE INVOLVED IN
 CC ANCHORING COMPONENTS OF THE PORE COMPLEX TO THE PORE MEMBRANE.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX. LOCATED TO THE
 CC TERMINAL RING STRUCTURE OF THE NUCLEOPLASMIC CAGE.
 CC -1- DOMAIN: CONTAINS X-F-X-F-G REPEATS.
 CC -1- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS
 CC SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST
 CC NUP1, NSP1, POM 121 AND MAMMALIAN P62.
 CC -1- SIMILARITY: CONTAINS 4 RANBP2-TYPE ZINC FINGERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: Z25535; CAA80982.1; -
 DR MIM: 603948; -
 DR InterPro: IPR001876; Znf-RanBP.
 DR Pfam: PF00641; Zf-RanBP; 4.
 DR SMART: SM00547; Znf_RBZ; 4.
 DR PROSITE: PS01358; ZF-RANBP2_1; 4.
 DR PROSITE: PS0199; ZF-RANBP2_2; 4.
 KW Nuclear protein; Transport; Repeat; Zinc-finger; DNA-binding.
 FT DOMAIN 4 14 POLY-GLY.
 FT ZN_FING 443 447
 FT ZN_FING 657 687 RANBP2-TYPE 1.
 FT ZN_FING 722 751 RANBP2-TYPE 2.
 FT ZN_FING 793 822 RANBP2-TYPE 3.
 FT ZN_FING 851 880 RANBP2-TYPE 4.
 SQ SEQUENCE 1475 AA; 153889 MW; 3CB415A6909DE80E CRC64;
 Query Match 7.7%; Score 96; DB 1; Length 1475;
 Best Local Similarity 27.5%; Pred. No. 5.1; Mismatches 47; Indels 30; Gaps 7;
 Matches 38; Conservative 23;
 OY 41 GAMVSTANFSGGTGAPATYPSYIKGCH-----WGNCYTKNKGMPDIQISQIG 88
 DB 1342 GSISSTALFPTGSPAPPTFTGVSSSSQPVFGQOPQSASRGSGTTPSSSAF---QFG 1398
 OY 89 SAVTSMTQVS-SCAYDYADIMTNSPTTTGGPNGEFTIMWLSRGCVQPFQSGQTATG 147
 DB 1399 SSTNENFTNNSPGVFTFGAN---SSTPMAASAPSGS-----GGF-PF-NQSPAA 1444
 OY 148 VTVAGTVMWMOGQOTSM 165
 DB 1445 FTVGSGKKNVSSSGTSF 1462

```

RESULT 13
YOR3_SOUV3 STANDARD; PRT: 211 AA.
ID 004550;
AC 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
OS Southampton virus (serotype 3).
OC Norwalk-like viruses.
NCBI_TaxID=37129;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93142023; PubMed=8380940;
RA Lambden P.R., Caul E.O., Ashley C.R., Clarke I.N.;
RT "Sequence and genome organization of a human small round-structured
(Norwalk-like) virus."
RL Science 259:516-519(1993).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
CC EMBL: L07418; AAA92985.1; -
DR PIR: C37491; C37491.
DR InterPro: IPR004278; RNA_capsid.
DR Pfam: PF03035; RNA_capsid.1.
KM Hypothetical protein.
SQ SEQUENCE 211 AA; 22257 MW; 3D6D1B51CC3C27B4 CRC64;

Query Match 7.6%; Score 95; DB 1; Length 211;
Best Local Similarity 26.8%; Pred. No. 0.75;
Matches 53; Conservative 18; Mismatches 57; Indels 70; Gaps 10;

QY 19 YRVQTNEMNSSAOQCLTINTATG---AMTVSTANFSGGTGADATYPSIKGCH---WG 71
Db 55 YQVQA-----SNALLAKNLNRYSMILAGLSADASRAVAGAPATRLIDNNGFVAPRS 110
QY 72 NCTTKNNG-----MPQIQSIGSAVTSMTQVSSGAVDAVDIWTNSPTTGGPNGT 125
Db 111 SATTLRSGGFMAVPMYQPR-----SKTPQSSGFSNPAYDMSTVSSRTSS----- 155
QY 126 EIMIML---NSRGGVOPFGSQTATGTVVAGHTWNVWQOOTSMKLIISYVLTGPTSTIS-- 180
Db 156 ----WQSQNSLSRVSFFHQAL-----QTWV-----VTPPGSTSSSV 190
QY 181 -----NLDLKAFAD 190
Db 191 SSTPYGVFNTDRMPLFAN 208

RESULT 14
MUR2_ENTHR STANDARD; PRT: 666 AA.
ID 004550;
AC 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Muramidase-2 precursor (PC 3.2.1.17) (1,4-beta-N-
acetyluramoylhydrolase) (Peptidoglycan hydrolase) (Pg-hydrolase-2)
(Lyxosyme).
OS Enterococcus hirae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
Enterococcus.
NCBI_TaxID=1354;

```

```

RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 50-73.
RC STRAIN-ATCC 9790;
RX MEDLINE=92165737; PubMed=1347040;
RA Chu C.-P., Kariyama R., Daneo-Moore L., Shockman G.D.;
RT "Cloning and sequence analysis of the muramidase-2 gene from
Enterococcus hirae."
RL J. Bacteriol. 174:1619-1625(1992).
RN [2]
RP FUNCTION.
RC STRAIN-ATCC 9790;
RX MEDLINE=89327152; PubMed=2753858;
RA Dolinger D.L., Daneo-Moore L., Shockman G.D.;
RT "The second peptidoglycan hydrolase of Streptococcus faecium ATCC 9790
covalently binds penicillin."
RL J. Bacteriol. 171:4355-4361(1989).
RN [3]
RP FUNCTION.
RC STRAIN-ATCC 9790;
RA Del Mar Lleo M., Canepari P., Satta G.;
RT "Thermosensitive cell growth mutants of Enterococcus hirae that
elongate at non-permissive temperature are stimulated to divide by
parental autolytic enzymes."
RL J. Gen. Microbiol. 139:3099-3117(1993).
CC -1- FUNCTION: MAY WORK IN CONCERT WITH AND POTENTIATE THE PROGRESSIVE
HYDROLYTIC ACTION OF MURAMIDASE-1, WHICH REQUIRES BINDING OF THE
ENZYME TO NONREDUCING ENDS OF GLYCAN CHAINS. HYDROLYSIS IN THE
MIDST OF GLYCAN CHAINS WOULD INCREASE THE NUMBER OF BINDING SITES
FOR MURAMIDASE-1. MAY FUNCTION IN FACILITATING SEPTUM FORMATION
AND CELL SEPARATION. ACTIVE ON M.LUTEUS CELL WALLS AND ON E.HIRAE
CELL WALL FRACTIONS, BUT NOT ACTIVE ON E.HIRAE INTACT CELL WALLS.
CC CAN COVALENTLY BIND PENICILLIN.
CC CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between
acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan
heteropolymers of the prokaryotes cell walls.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: LYXN REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOLYCAN
BINDING.
CC -1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
CC -1- SIMILARITY: CONTAINS 6 LYXN REPEATS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
DR EMBL: M77639; AAA24776.1; -
DR PIR: A42296; A42296.
DR HSSP: P21697; 1PCS.
DR InterPro: IPR002901; Amidase_4.
DR Pfam: PF01832; Amidase_4; 1.
DR Pfam: PF01476; LyxM; 6.
DR SMART: SM00047; Lyx2; 1.
DR SMART: SM00257; LyxM; 6.
KW Hydrolase; Glycosidase; Bacteriolytic enzyme; Cell wall;
KW Cell division; Septation; Repeat; Signal.
FT SIGNAL 1 49
FT CHAIN 50 666 MURAMIDASE-2.
FT REPEAT 257 299 LYXN 1.
FT REPEAT 338 380 LYXN 2.
FT REPEAT 414 456 LYXN 3.
FT REPEAT 489 531 LYXN 4.
FT REPEAT 565 607 LYXN 5.
FT REPEAT 623 665 LYXN 6.
SQ SEQUENCE 666 AA; 70670 MW; FFOA7FACBD810BA3 CRC64;

```

Query Match 7.6%; Score 95; DB 1; Length 666;
Best Local Similarity 24.7%; Pred. No. 2.6;

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:37:51 ; Search time 125.36 Seconds
(without alignments)
318.777 Million cell updates/sec

Title: US-09-917-384-7

Perfect score: 1244
Sequence: 1 DCPGPNGVTSVQDEYR.....IMGGGGIGSNSFSVSTSG 231

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_19:*
1: sp_Archea:*
2: sp_Bacteria:*
3: sp_Fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_Organella:*
9: sp_Phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	472.5	38.0	371	2	09KIH1
2	449.5	36.1	377	2	008468
3	436.5	35.1	382	2	059963
4	434.5	34.9	384	2	09X602
5	425.5	34.2	381	2	09RJY3
6	424.5	34.1	381	2	054331
7	407.5	32.8	260	2	033897
8	235.5	18.9	151	16	053438
9	214.5	17.2	264	2	031030
10	208	16.7	263	2	09AN41
11	207.5	16.7	244	2	09XYS4
12	193.5	15.6	239	3	013454
13	192.5	15.5	239	3	074705
14	175.5	14.1	234	3	000095
15	175.5	14.1	238	3	094218
16	150	12.1	239	3	09P8N6

17	133	10.7	334	17	097X08	097X08 sulfolobus
18	128	10.3	332	17	097Y67	097Y67 sulfolobus
19	109	8.8	402	10	09AXB0	09AXB0 oryza sativ
20	108	8.7	407	3	000895	000895 collettotric
21	108	8.7	1275	5	076602	076602 caenorhabdl
22	108	8.7	1844	5	022579	022579 caenorhabdl
23	107.5	8.6	319	1	09V2T0	09V2T0 pyrococcus
24	107	8.6	392	10	096487	096487 lycopersico
25	106	8.5	780	2	054229	054229 streptomyce
26	105.5	8.5	257	2	060032	060032 thermotoga
27	105.5	8.5	258	16	09S5X8	09S5X8 thermotoga
28	105.5	8.5	958	16	09PBA7	09PBA7 xylella tas
29	105	8.4	461	2	059922	059922 streptomyce
30	100.5	8.1	464	9	09A286	09A286 lactobacill
31	100.5	8.1	1203	5	09N5K0	09N5K0 caenorhabdl
32	100	8.0	4776	16	097P71	097P71 streptococ
33	99.5	8.0	267	15	09WDJ9	09WDJ9 human immun
34	99.5	8.0	267	15	09WDK6	09WDK6 human immun
35	99	8.0	267	15	09WEA0	09WEA0 human immun
36	99	8.0	335	2	09ROB8	09ROB8 cellulomona
37	99	8.0	427	2	09FC07	09FC07 pseudomonas
38	98.5	7.9	267	15	09W9X0	09W9X0 human immun
39	98.5	7.9	526	11	063850	063850 mus sp. nuc
40	98.5	7.9	3930	16	09B220	09B220 rhizobium l
41	98	7.9	795	10	09LJ20	09LJ20 oryza sativ
42	98	7.9	1390	5	017602	017602 caenorhabdl
43	98	7.9	1749	16	P73032	P73032 synechocyst
44	97.5	7.8	1160	11	09WUN1	09WUN1 mus musculu
45	97.5	7.8	1198	11	099PM6	099PM6 mus musculu

ALIGNMENTS

RESULT 1
ID 09KIH1 PRELIMINARY; PRT: 371 AA.
AC 09KIH1;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELLDLASE 12A.
GN CEL12A.
OS Streptomyces sp. 11A68.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomyicinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=133452;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=11A68;
RA van Solingen P., Meljer D., van der Kleij W.A.H., Barnett C.C.,
RA Bolle R., Power S.D., Jones B.E.;
RT "Cloning and expression of an endocellulase gene from a novel
RT Streptomyces isolated from an East African soda lake."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF233376; AAF91283.1; -
DR HSSP: P07986; 1EXG.
DR InterPro: IPR002594; Glyco_hydro_12.
DR InterPro: IPR001230; Prenyltn.
DR Pfam: PF01670; Glyco_hydro_12; 1.
DR ProDom: PD004316; Glyco_hydro_12; 1.
DR PROSITE: PS00294; PRENYLATIOn; UNKNOWN.1.
SQ SQUENCE 371 AA; 38481 MW; 0E1BC4288A148914 CRC64;

Query Match 38.0%; Score 472.5; DB 2; Length 371;
Best Local Similarity 46.6%; Pred. No. 3.4e-27;
Matches 104; Conservative 28; Mismatches 82; Indels 9; Gaps 8;

QY 10 GVSVDGCEYRVQVQENMSSNAOCLTINTATGATVSTANFSGGTGAPATPISYIGCH 69
DB 40 GTTTID-RVYVQNNRGTSATOC--INTVNGEFTIQADGSVFTNGAPRSPSYVDGCH 96

```

OY 70 WNCNCTKNNKNGAIOISQIGSAVTSMTSQOVSSGAVDAVDADWTNSTPTTGGPNCGETEIM 129
      :|||:::||:|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 97 YENCAPRTT-LEMRISISSASPSYSRYRTGTGVYNAAYDILWDPPTPTNG-VNTEIEMT 154
OY 130 WLNSRGVOPEFGSQTATGTVTAHTVMWMOGOOTSMMKIISYLTFPGANSIMLDKAIFA 189
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155 WENRVPQVPIGPSPYGT-AHVGGRSMEWETGSNGSNDVISF-LAPSALISWSFEVKD-FV 211
OY 190 DAAAGSLNTSD-YLLDVEAGFEINOGGCGLSNSEFSVSYS 231
Db 212 DQAVSHGLATPDWYLTLSIOAGFEPMEGGTGTLVNSFFSSAVNMG 254

RESULT 2
008468 PRELIMINARY: PRT: 377 AA.
AC 008468;
DR 01-JUL-1997 (TREMBLrel_04, Created)
DT 01-JUL-1997 (TREMBLrel_04, Last sequence update)
DE 01-DEC-2001 (TREMBLrel_19, Last annotation update)
DE CEL2 (EC 3.2.1.4).
GN CELA2.
OS Streptomyces halstedii.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_Taxid=1944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JM8;
RX MEDLINE=97307849; PubMed=9182697;
RA Garcia-Salas A.L., Fernandez-Albalos J.M., Sanchez P., Ruiz-Arribas A.,
RA Santamaria Sanchez R.I.;
RT "Two genes encoding an endoglucanase and a cellulose-binding protein
RT are clustered and co-regulated by a TTA codon in Streptomyces
RT halstedii JM8." ;
RL Biochem. J. 324:403-411(1997).
DR EMBL: U51222; AAC45429.1; -.
DR HSRP; F07986; IEXG.
DR InterPro: IPR001919; CBD_2.
DR InterPro: IPR002594; Glyco_hydro_12.
DR Pfam: PF00553; CBD_2; 1.
DR Pfam: PF01670; Glyco_hydro_12; 1.
DR Prodom: PD004316; Glyco_hydro_12; 1.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 377 AA; 38829 MW; 9892191B741EC280 CRC64;

Query Match 36.1%; Score 449.5; DB 2; Length 377;
Best Local Similarity 43.8%; Pred. No. 1.7e-25;
Matches 98; Conservative 31; Mismatches 88; Indels 7; Gaps 6;

OY 8 QNGVTSVODGEFRVOTNMENWSAQCCLINTATGAWTSTANFSGGTGAPATPYSIKG 67
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 44 QVGSITLIQ-GRVVQNRRKGASAPQCVT--AIDSGFRVQAQALCAPTNGAPKSPSVENG 100
OY 68 CHWGNCCTTKNVAMPILQISOISGAVTSWSTQVSSGAYDVADYIMTNSFTTTGGPNGTEI 127
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 101 CHTYTC-SPGTMLPAOVSGIASAPASISYGFVGSAYVNASTYDIWDLPFKKNG-VNRTEI 158
OY 128 MTWLMSRGVOPFGSGOTANGVTVAAGTHMWVMOGOOTSMMKIISYLTPGATSISNLDKAI 187
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 159 MTWLNKVGIQIPGISQAGT-ASVGGRTQVWMVNGSNGSNDVISFV-APSAVASWSFDYDMF 216
OY 188 FDAAARGSLNTSDYLLDVEAGFEINOGGCGLSNSEFSVSYS 231
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 217 VANTARAGMAONMYLTISVQACFEFMONGAGLVNSSFSTVNLG 260

RESULT 3
059963 PRELIMINARY: PRT: 382 AA.
AC 059963;
DC 059963;
DT 01-NOV-1996 (TREMBLrel_01, Created)
```

DT	01-NOV-1996 (TReMBLrel. 01, last sequence update)
DT	01-DEC-2001 (TReMBLrel. 19, last annotation update)
DE	CELLULOSE (EC 3.2.1.4).
GN	EGIS.
OS	Streptomyces rochei (Streptomyces parvullus).
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
CC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX	NCBI_TaxID=1928;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-A2;
RX	MEDLINE=95011642; PubMed=7523249;
RA	Perito B., Hanhart E., Irdani T., Iqbal M., McCarthy A.J.,
RA	Mastromeli G.,
RT	"Characterization and sequence analysis of a Streptomyces rochei A2,"
RL	Gene 148:119-124(1994).
DR	EMBL: X73953; CAAS2139.1; "
DR	HSSP: P07986; IEXG.
DR	InterPro: IPR001919; CBD_2.
DR	InterPro: IPR002594; Glyco_hydro_12.
DR	Pfam: PF00553; CBD_2; 1.
DR	Pfam: PF01670; Glyco_hydro_12; 1.
DR	ProDom: PD04316; Glyco_hydro_12; 1.
DR	ProSite, PS00561; CBD_BACTERIAL; UNKNOWN_1.
KW	Hydrolase; Glycosidase.
SO	SEQUENCE 382 AA; 39398 MW; 21C014342EFC6565 CRC64;
Query Match	35.1%; Score 436.5; DB 2; Length 382;
Best Local Similarity	43.2%; Pred. No. 1.6e-24;
Matches	96; Conservative 30; Mismatches 89; Indels 7; Gaps 6;
QY	10 GVTSDGEYVOTENMNSAQCCLITFAGAMVTSTANSGGTGAPATPSITYGCH 69
Db	46 GSTVIQ-GRYVQNNRMWTSATQCVT--ATDSGRVYQADSDSVTLNGAPKPSVFNCGH 102
QY	70 WGNCTTKNVGMPDIQISQISAVTSMSTTQVSSGAYDAVDIMNSTPTTGGPNGTEIMI 129
Db	103 YTNCG-SPTGALPARISGISAPSSISYGFVDNAVYNSYDIDLPDPTPDG-VNRETEIMI 160
QY	130 WLNSSGVQPPRGSGQTATGVTVAAGHTHWNWQOQTSWKIISVLTLPGATISNDLKAIFA 189
Db	161 WFNRYGQIQPIGSGOQGT-ASVAGRWTEWSSGNGTNDLSFV-APSAWSSWSPVMDFVR 218
QY	190 DAAAGSLNTSDYLDLVDAGEFEINOGGOGGLSGNSFSVSTG 231
Db	219 ATVARGLAGNDMYLTLSIQAGFEPPWNGAGLAVNSFSSTVNTG 260

DR InterPro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF00553; CBD_2; 1.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR ProDom: PD004316; Glyco_hydro_12; 1.
 DR PROSITE: PS00561; CBD_BACTERIAL; UNKNOWN_1.
 SQ SEQUENCE 384 AA; 40918 MW; D3968BFB6BDEDE65 CRC64;

Query Match 34.9%; Score 434.5; DB 2; Length 384;
 Best Local Similarity 42.7%; Pred. No. 2, 2e-24;
 Matches 102; Conservative 28; Mismatches 90; Indels 19; Gaps 9;

QY 2 CTRGPONGVTSVQDEYRQTMENSSAOCLITNTATGAMTVSTANFSGGTGAPATY 61
 DB 42 CEP-----YGTITIQ-GRVYQNNRKGSSPOCVTA-TDTG-FRLTQADGVPPNGAFKSY 94
 QY 62 PSIRKCGHMGNCNT--TK-----NVMGPIQISGSAVTSWSTOVSSGAYDAVDIWT 112
 DB 95 PSYFNCGHTNCGSPGTRKLPARIGISAPARIGISSAPSSISYGVGAVYNASDIWL 134
 QY 113 NSTPTTGGPNGTEIMIMLNSRGVQPGSGQTATGTVAGHTNWNWOGQOTSMKIIISYVL 172
 DB 155 DPTPRPDG-VNRTEIMTFKRVGPIDPISQVGT-ATVGRTWQVNSGNGSNDVLSFV- 211
 QY 173 TREATSISNLDKAFADAAARSLNTSDYLLDVEAGFEIMOGGCGSGNSFSVSTSG 231
 DB 212 APSAIESWSEFDVMDPEVETVARGMAQNDWYLTISVQAGFEPMONGAGLAVNSFSSVDTVG 270

RESULT 5

Q9RJY3 PRELIMINARY; PRT; 381 AA.

AC Q9RJY3; 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PUTATIVE SECRETED CELLULOSE B.
 GN CELB.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Oliver K., Harris D.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Cerdano A.M., Parkhill J., Bartell B.G., Rajandream M.A.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kleiser H.M., Denaplatte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RT Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL133210; CAB61599.1; -;
 DR HSSP: P07986; 1EXG.
 DR InterPro: IPR001919; CBD_2.
 DR InterPro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF00553; CBD_2; 1.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR ProDom: PD004316; Glyco_hydro_12; 1.
 DR PROSITE: PS00561; CBD_BACTERIAL; UNKNOWN_1.
 SQ SEQUENCE 381 AA; 39199 MW; 97CD8FF58679E4EC CRC64;

Query Match 34.2%; Score 425.5; DB 2; Length 381;
 Best Local Similarity 42.3%; Pred. No. 1e-23;

Matches 94; Conservative 32; Mismatches 89; Indels 7; Gaps 7;

QY 10 GYTSVQDEYRQTMENSSAOCLITNTATGAMTVSTANFSGGTGAPATYPSIRKCGH 69
 DB 49 GTTITIQ-GRVYQNNRKGSSPOCVTA-TDTG-FRLTQADGVPPNGAFKSY 105
 QY 70 WGNCTTKNMGPIQISGSAVTSWSTOVSSGAYDAVDIWTNSTPTTGGPNGTEIMI 129
 DB 106 YTNCSGTPALPRLDVTVAAPSISYGFVDGAVYNASDIWLDPTARTDG-VNQTETMI 163
 QY 130 WLSNRGVQPGSGQTATGTVAGHTNWNWOGQOTSMKIIISYVLPLGATSIISNLDKAFIFA 189
 DB 164 WFNRRVGPIDPISQVGT-ASVGRTWQVNSGNGSNDVLSFV-APSAISGMSFDVMDPVR 221
 QY 190 DAARGSLNTSDYLLDVEAGFEIMOGGCGSGNSFSVSTSG 231
 DB 222 ATYARGLAENDWYLTISVQAGFEPMONGAGLAVNSFSSVDTVG 263

RESULT 6

O54331 PRELIMINARY; PRT; 381 AA.

AC O54331; 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CELLULOSE B PRECURSOR.
 GN CELB.
 OS Streptomyces lividans.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1916;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-66;
 RA Shareck F.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U04629; AAB71950.1; -;
 DR HSSP: P07986; 1EXG.
 DR InterPro: IPR001919; CBD_2.
 DR InterPro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF00553; CBD_2; 1.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR ProDom: PD004316; Glyco_hydro_12; 1.
 DR PROSITE: PS00561; CBD_BACTERIAL; UNKNOWN_1.
 KW SIGNAL.
 FT SIGNAL 1 40 POTENTIAL.
 FT CHAIN 41 381 CELLULOSE B.
 SQ SEQUENCE 381 AA; 39239 MW; A7E99BF590A24EC CRC64;

Query Match 34.1%; Score 424.5; DB 2; Length 381;
 Best Local Similarity 42.3%; Pred. No. 1, 2e-23;
 Matches 94; Conservative 32; Mismatches 89; Indels 7; Gaps 7;

QY 10 GYTSVQDEYRQTMENSSAOCLITNTATGAMTVSTANFSGGTGAPATYPSIRKCGH 69
 DB 49 GTTITIQ-GRVYQNNRKGSSPOCVTA-TDTG-FRLTQADGVPPNGAFKSY 105
 QY 70 WGNCTTKNMGPIQISGSAVTSWSTOVSSGAYDAVDIWTNSTPTTGGPNGTEIMI 129
 DB 106 YTNCSGTPALPRLDVTVAAPSISYGFVDGAVYNASDIWLDPTARTDG-VNQTETMI 163
 QY 130 WLSNRGVQPGSGQTATGTVAGHTNWNWOGQOTSMKIIISYVLPLGATSIISNLDKAFIFA 189

Matches 66; Conservative 38; Mismatches 98; Indels 29; Gaps 11;

QY 15 QDEYRVQTEEMNSA-----OQCLTINTATGA-----WTVSTANFGSGTGAPATPSIX 65
 DB 45 ENKRYTFNNWVGKDEKMGQGYFVNSPTSMGNMHPSSSSVK-----AYSLV 96
 QY 66 KCGHMGCTTKNVGMPIQISQISGAVTSWSTQVSSGA-YDAVDIWTNSTPTTGPNG 124
 DB 97 SGWHMTAVYTEENSGLPKILSS-NKSIYSNTYYSIKSTATILYAAYDVWFHTDKASMDSTP 155
 QY 125 T-ETIMILNLRGCGVQPGSGQTATGVTAGHTWNNWQ-----GQOTSMKIISYLTLPGATS 178
 DB 156 TDELMTMLNMT-NAGPAGDIET-VFLGSSMWFYKGMINAGGNNVSEFVTSNTNS 213
 QY 179 ISNLDLKAIFADAAARCS-INTSDYLDVDEAGFEIMOGGCGIGNSFSVS 228
 DB 214 AS-LNIHFINDVGRRKMSNAKISSVELGTETIFGCDQIDITKMSVDY 263

RESULT 10
 Q9AN41 PRELIMINARY; PRT; 263 AA.
 AC 09AN41;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
 DE ID568.
 GN 1D568.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Bradyrhizobium group; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=110594;
 RC MEDLINE=21101824; PubMed=11157954;
 RA Gottfert M., Rothlisberger S., Kundig C., Beck C., Marty R.,
 RA Hennecke H.;
 RT "Potential symbiosis-specific genes uncovered by sequencing a 410-kb
 RT DNA region of the Bradyrhizobium japonicum chromosome.";
 RT J. Bacteriol. 183:1405-1412(2001).
 DR EMBL: AF322013; MAG60937.1;
 DR InterPro: IPR002594; Glyco_hydro_12.
 DR ProDom: PD004316; Glyco_hydro_12; 1.
 SQ SEQUENCE 263 AA; 28509 MW; 8F66BAF9239487D6 CRC64;

Query Match 16.7%; Score 208; DB 2; Length 263;
 Best Local Similarity 28.2%; Pred. No. 6.1e-08;
 Matches 69; Conservative 43; Mismatches 75; Indels 58; Gaps 14;

QY 12 TSVDGEYRVQTEEMNSA-----SAOCLTINTAT--GAWTVSTANFGSGTGAPATPSIX 63
 DB 50 SSAOYGFSDIGYSWNNNDVWGRGAPQTISAVANQGVWSNQP-----DTGGI-KSTP- 102
 QY 64 IYKCHMGCTTKNVGMPIQISQISGAVTSWSTQVSSGAYDAVDIWTNSTPTTGPNG 123
 DB 103 -----HGG-----FMVGRP--LSSINTLISNNGEVPISGANDVAYDIDMS-----N 143
 QY 124 GTEIMILN-----SRGCVQPGSGQA-----TGTVAGHTWNNWQOQTSKII 168
 DB 144 QYEIMLTNTYTGNDGCGNKPISYKYASSPAIPITYNVDVGATNNVFG-ENHHRVI 202
 QY 169 SYVLTLPATISNLDLKAIFADAAR--GSLNTSDYLDVDEAGFEIMOGGCGIGNSFS 225
 DB 203 SFLMT-SKTSNGTIDKSVLQWIKSKYFGEINVG-----NVQGEVLTSSPGMNFENN 257
 QY 226 VSVTS 230
 DB 258 WTLTS 262

RESULT 11

Q9KTS4
 ID 09KTS4 PRELIMINARY; PRT; 244 AA.
 AC 09KTS4;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE PUTATIVE SECRETED SUGAR HYDROLASE.
 GN SC5H4.15.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Murphy L., Harris D.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RC MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kleser H.M., Denaparte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmid and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RT Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL355913; CAB91127.1;
 DR InterPro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR ProDom: PD004316; Glyco_hydro_12; 1.
 KW Hydrolase.
 SQ SEQUENCE 244 AA; 26123 MW; 2CA63242798EAD4 CRC64;

Query Match 16.7%; Score 207.5; DB 2; Length 244;
 Best Local Similarity 28.3%; Pred. No. 6.1e-08;
 Matches 65; Conservative 44; Mismatches 84; Indels 37; Gaps 12;

QY 7 NQNGTSVQDEYRVQTEEMNSA-----OQCLTINTAT--GAWTVSTANFGSGTGAPATPSIX 63
 DB 47 DQNGTSLNG--YTLXNNIMSGCAGSCVYANSGTDWV--ADHNTDG-----IKSTPN 98
 QY 64 IYKCHMGCTTKNVGMPIQISQISGAVTSWSTQVSSGAYDAVDIWTNSTPTTGPNG 123
 DB 99 -----AKVIYIKP--ITSLSLTSSYNTVTPSSGAYNTSYDIDTD-----Y 138
 QY 124 GTEIMILNFGCVQPGSGQTATGVTAGHTWNNWQOQTSKIIISYLTLPGATS--ISN 181
 DB 139 DYEIMLWNHGAAGPLDTFGS--VGLGHTWVDYKGNNGANEVFSFRTSDNSGTANI 197
 QY 182 LDLKAIFADAARSLNTSDYLDVDEAGFEIMOGGCGIGNSFSVS 231
 DB 198 LPIKWIWD--TKGMNG-NETIGDVQFYEITSSAGIGLDFRTNLTGSG 244

RESULT 12
 ID 013454 PRELIMINARY; PRT; 239 AA.
 AC 013454;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4).
 GN CELA.
 OS Aspergillus oryzae.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillius.
 OX NCBI_TaxID=5062;

[1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-KBN616;
 RA MEDLINE-97161783; PubMed-9008887;
 RA Kitamoto N., Go M., Shibayama T., Kimura T., Kito Y., Ohmura K.,
 RA Tsukagoshi N.;
 RT "Molecular cloning, purification and characterization of two endo-1,4-
 RT beta-glucanases from *Aspergillus oryzae* KBN616."
 RL Appl. Microbiol. Biotechnol. 46:538-544(1996).
 DR Interpro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR Prodom: PD004316; Glyco_hydro_12; 1.
 KW Hydrolase; Glycosidase
 SQ SEQUENCE 239 AA; 26096 MW; C0F850E5DFE455D CRC64;

Query Match 15.6%; Score 193.5; DB 3; Length 239;
 Best Local Similarity 27.4%; Pred. No. 6.3e-07;
 Matches 60; Conservative 39; Mismatches 87; Indels 33; Gaps 10;

QY 13 SVODGEYRVQTNW--NSSAOQCLTIN--TATGAMTVSTANFSGTGAPATYPSIYNG 67
 DB 25 SASPPYSVNNMLMGODSGCYVDNLSSGAAMHTTWTMNGEGSVKS-----76
 QY 68 CHMGCTKKNVGMPIQISGSAVTSMTQVSSGAYDAVDIWTNSTPTTGGPNGTEI 127
 DB 77 -YSNNAVTFDKLVSDVOSIPTDV--EWSODNTNVA-DVAYDLFTAAADONHTVYSGDYEL 133
 QY 128 MIMLNSRGVQPFSGQTATGTVAGHTMNVWQ-----QOTSWKIISYVLTPGATIS 180
 DB 134 MIMLNRGVIQPIGQIDP-ATVEGHTWELWGTITIGAGAEKRTSFFV-----ATPIPN 186

QY 181 NL--DKAIFADAAAGSLNLS-DYLLDVEAGFEIWMQ 216
 DB 187 TFGDIKKEFFDYITSKHSPASQAQYLINMQFTPEPTGG 225

RESULT 13
 ID 074705 PRELIMINARY; PRT; 239 AA.
 AC 074705;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ENDOGLUCANASE A PRECURSOR (EC 3.2.1.4) (CELLULOSE)
 DE (ENDO-1,4-BETA-GLUCANASE A) (CARBOXYMETHYLCELLULOSE).
 GN EGIA OR CEKA.
 OS *Aspergillus niger*.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; *Aspergillus*.
 OX NCBI_TaxID=5061;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CBS 120.49 / N400;
 RX MEDLINE-98432774; PubMed-9758775;
 RA van Peijl N.M.M.E., Gietkens M.M.C., de Vries R.P., Visser J.,
 RA De Graaf L.H.;
 RT "The transcriptional activator XlnR regulates both xyloanalytic and
 RT endoglucanase gene expression in *Aspergillus niger*."
 RL Appl. Environ. Microbiol. 64:3615-3619(1998).
 CC -1- FUNCTION: HAS CARBOXYMETHYLCELLULOSE ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY H (FAMILY 12 OF GLYCOSYL
 CC HYDROLASES).
 DR EMBL: AJ224451; CA11964.1; -;
 DR Interpro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR Prodom: PD004316; Glyco_hydro_12; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 16 POTENTIAL.
 FT CHAIN 17 239 ENDOGLUCANASE A.

SQ SEQUENCE 239 AA; 25859 MW; 5EB6267DA54AA2A CRC64;

Query Match 15.5%; Score 192.5; DB 3; Length 239;
 Best Local Similarity 28.8%; Pred. No. 7.5e-07;
 Matches 63; Conservative 33; Mismatches 90; Indels 33; Gaps 11;

QY 13 SVODGEYRVQTNW--NSSAOQCLTIN--TATGA-----WTVSTANFSGTGAPATYP 62
 DB 25 SASPPYSVNNMLMGEOGTGSCYVDNLSSGASMHTEWT-----MGEGEYVKS---76
 QY 63 SYKGGHMCNCTKKNVGMPIQISGSAVTSMTQVSSGAYDAVDIWTNSTPTTGGP 122
 DB 77 -----YSNNAVTFDKLVSDVSSIPTSV--EMKQDNTNVA-DVAYDLFTAAANDHATSS 128

QY 123 NCTEIMIMLNSRGVQPFSGQTATGTVAGHTMNVWQGOOT---SWKIISYVLTPGATIS 178
 DB 129 GYELMIMLNRGVIQPIGQIDP-ATVGGKMEVWYSGTTQAGAEQRTSYVESPIINS 187

QY 179 ISNLDKAIADAA--RGSLSNTSDYLLDVEAGFEIWMQ 216
 DB 188 YSG-DINAFPSYLTQNGFPASSQYDLINQFTEAFTGG 225

RESULT 14
 ID 000095 PRELIMINARY; PRT; 234 AA.
 AC 000095;
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE ENDO-BETA-1,4-GLUCANASE (EC 3.2.1.4).
 GN EGL.
 OS *Trichoderma reesei* (Hypocrea jecorina).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; Hypocreaceae; *Hypocrea*.
 OX NCBI_TaxID=51453;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-QM9414;
 RA Okada H., Tada K., Sekiya T., Yokoyama K., Takahashi A., Tonda H.,
 RA Kumagai H., Morikawa Y.;
 RT "Molecular characterization and heterologous expression of the gene
 RT encoding a low-molecular mass endoglucanase from *Trichoderma reesei*
 RT QM9414."
 RL Appl. Environ. Microbiol. 64:55-563(1998).
 DR EMBL: AB003694; BAA20140.1; -;
 DR Interpro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR Prodom: PD004316; Glyco_hydro_12; 1.
 KW Hydrolase; Glycosidase.
 SQ SEQUENCE 234 AA; 25159 MW; DF476EDE384ADD1 CRC64;

Query Match 14.1%; Score 175.5; DB 3; Length 234;
 Best Local Similarity 26.7%; Pred. No. 1.3e-05;
 Matches 60; Conservative 34; Mismatches 94; Indels 37; Gaps 9;

QY 19 YVQTNENNSSAOQCLTINTATGAMTVSTANFSGTGAPATYPSIYKCHMGNCCTKNV 78
 DB 31 YVSNMIMKASA-----GSGFCYTAIVSLSGA-----SMHADWQNSG--GQNN 72

QY 79 GMPDIQISQIG-----SAVTSMTQVSSGAY-----DVAYDIWTNSTPTTGGPNG 124
 DB 73 VRSYQNSQIAIPQKRTVNSISSMPTT--ASWISYSGSNIRANAYDLFTAAANHTVYSGD 130

QY 125 TEIMIMLNSRGVQPFSGQTATGTVAGHTMNVWQGOOTSMKIISYVLTPGATISINDL 184
 DB 131 YELMIMLNRGVIQPIGQIDP-ATVGGKMEVWYSGTTQAGAEQRTSYVESPIINS 187

QY 185 KAIIFADAAAGSLNLS-DYLLDVEAGFEIWMQGOIGLGSFVSV 228
 DB 189 KNFENLNRKNKGYNMAAGYVLSYQFTPEPTGSGTILNVAWSWTASI 233

RESULT 15

094218 PRELIMINARY; PRT; 238 AA.
 ID 094218
 AC 094218;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE XYLOGLUCAN-SPECIFIC ENDO-BETA-1,4-GLUCANASE PRECURSOR.
 OS Aspergillus aculeatus.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxId=5053;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-KSM 510;
 RX MEDLINE=99102417; PubMed=9884411;
 RA Pauly M., Andersen L.N., Kauppinen S., Kofoed L.V., York W.S.,
 RA Albersheim P., Darvill A.;
 RT "A xyloglucan-specific endo-beta-1,4-glucanase from Aspergillus
 RT aculeatus: expression cloning in yeast, purification and
 RT characterization of the recombinant enzyme."
 RL Glycobiology 9:93-100(1999).
 DR EMBL; AF043595; AAD02275.1;
 DR InterPro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR ProDom: PD004316; Glyco_hydro_12; 1.
 KW Signal.
 FT SIGNAL 1 14 POTENTIAL.
 FT CHAIN 15 238 XYLOGLUCAN-SPECIFIC ENDO-BETA-1,4-
 FT GLUCANASE.
 SQ SEQUENCE 238 AA; 25158 MW; FCCA6746D9AEC1B1 CRC64;

Query Match 14.1%; Score 175.5; DB 3; Length 238;
 Best Local Similarity 26.8%; Pred. No. 1.3e-05;
 Matches 61; Conservative 35; Mismatches 93; Indels 39; Gaps 10;

QY 17 GEYRVQTNENWSSA--OQCLTINTANG--AWTVSTANFSGGNG-----GAPATYPSI 64
 DB 33 GDTLYNDLMGESAGTGSQCTGYDSISGDTIANHTSW-SMSGSSSVKSYVNAALTF--- 88
 QY 65 YKCGHMGNCITKNVGMPIQISQIGSAVTSWS-TTOVSSGAYDYADYDIWNSPTTGQPN 123
 DB 89 -----TPQLNCISSTIPTWKWYSGSSIVADYADYDFLEAFASGSSK-- 131
 QY 124 GTIIMITLNSRGCVQPPGS--QTATGYTVAGHTWVNWQGOQTSNKIISYVLPQATISIN 181
 DB 132 -YEIMVWLAALGAGPISSTGSTIATPTIAGVWVKLYSGPNCDDTYYSFVADSTTESFSG 190
 QY 182 LDKAIFADAAARGSLNTSDYLDVEAGFEIMOGGCG-LGSNSPSVSV 228
 DB 191 -DLNDEFTYIYLVNREGVSDELYLTLEAGTEPFGSNKLTVEYSYISI 237

Search completed: August 29, 2002, 16:37:53
 Job time: 963 sec

This Page Blank (uspto)

Db 112 YGNCAPPTT-LPMRISISGAPSSSVRYTNGVYNAAYIMWDPRTNG-VNRTIIMI 169
QY 130 WLSRSGVOPFGSQTATGVVAGHTWVWMOGQOTSMKIIISYVLTGPATISINLKAIFA 189
Db 170 WFRNRPVOPIGSPVGT-AHVGRSMWMTGSGNSNDVLSFV-LAPSAISSMSPVKD-FV 226
QY 190 DAARGLSNTSD-YLDDVEAGFEIMOGGGLGNSFSVSYSVTG 231
Db 227 DQAVSHGLATPDMYLTLSIQAGFEPMEGTGLAVNSFSSAVNAG 269

RESULT 5
US-09-216-295-22
; Sequence 22, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Wendt, Dan J.
; TITLE OF INVENTION: No. 6268328e1 Variant EGIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22
; LENGTH: 381
; TYPE: PRF
; ORGANISM: Streptomyces lividans CelB
US-09-216-295-22

Query Match 34.1%; Score 424.5; DB 4; Length 381;
Best Local Similarity 42.3%; Pred. No. 7e-32; Mismatches 89; Indels 7; Gaps 7;
Matches 94; Conservative 32;

QY 10 GVTSGVQDEYVQVTNEMSSAQOCLTINTGAMTVSTANFSGGTGAPATPYSYKGC 69
Db 49 GTTIIQ-GRYVONNRKSGTAPQCVR-TDTG-FRVTQADGSAPTNCAFKSPSEVFGCH 105
QY 70 WGNCTKKNVGMPIQISIGSAVTSMTQVSSGAYDAVIMTNSPTTGGPNGTEIMI 129
Db 106 YTNCG-SPGTDLPVRLDVSAAPSSISYGFVDAVYASVIMLDPARTDG-VNQTETIMI 163
QY 130 WLSRSGVOPFGSQTATGVVAGHTWVWMOGQOTSMKIIISYVLTGPATISINLKAIFA 189
Db 164 WFRNRPVOPIGSPVGT-AHVGRSMWMTGSGNSNDVLSFV-LAPSAISSMSPVKD-FV 221
QY 190 DAARGLSNTSD-YLDDVEAGFEIMOGGGLGNSFSVSYSVTG 231
Db 222 ATVARGLAENDMYLTLSVQAGFEPMONGAGLAVNSFSSSTVETG 263

RESULT 6
US-09-216-295-23
; Sequence 23, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Wendt, Dan J.
; TITLE OF INVENTION: No. 6268328e1 Variant EGIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 23
; LENGTH: 260
; TYPE: PRF
; ORGANISM: Rhodothermus marinus
US-09-216-295-23

Query Match 32.8%; Score 407.5; DB 4; Length 260;
Best Local Similarity 40.1%; Pred. No. 1.6e-30;
Matches 89; Conservative 30; Mismatches 84; Indels 19; Gaps 7;
QY 14 VOGEYVQVQNMENSSAQOCLTITATGAMTVSTANFSGGTGAPATPYSYKCHWG-- 71
Db 50 VAGGRYVQVNNWGAETAOCEVLETFETITPADHNGNN--VAAYPALYFEGCHWAPA 107
QY 72 ----NCTTKNVGMPIQISIGSAVTSMTQVSSGAYDAVIMTNS-PTTGGPNGTE 126
Db 108 RAIRDCAR-AGAVRAHEL-----DYTPITGRNNAAYDIFSPYTSNGKYSGAE 159
QY 127 IMTWLSRSGVOPFGSQTATGVVAGHTWVWMOGQOTSMKIIISYVLTGPATISINLKAIFA 186
Db 160 IMTWLSRSGVOPFGSQTATGVVAGHTWVWMOGQOTSMKIIISYVLTGPATISINLKAIFA 217
QY 187 IFADAAAGSLNTSD-YLDDVEAGFEIMOGGGLGNSFSVS 228
Db 218 FIDDAVARGYIRPEMYLHAAVETGELWEGAGLTADEFSTV 259

RESULT 7
US-09-216-295-5
; Sequence 5, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Wendt, Dan J.
; TITLE OF INVENTION: No. 6268328e1 Variant EGIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 259
; TYPE: PRF
; ORGANISM: Aspergillus aculeatus
US-09-216-295-5

Query Match 20.9%; Score 260.5; DB 4; Length 259;
Best Local Similarity 33.9%; Pred. No. 8.6e-17;
Matches 82; Conservative 21; Mismatches 96; Indels 43; Gaps 10;

QY 17 GEYVQVQNMENSSAQOCLTITATGAMTV-STANFSGGTGAPATPYSYKCHWG 71
Db 31 GYTTINNNLNGKDGAGSGSOCTTVNSASAGTSMTKKNWSGGENSVKS-----YAN 81
QY 72 NCTTKNVGMPIQISIGSAVTSMTQVSSGAYDAVIMTNSPTTGGPNGTEIMI 131
Db 82 SGLTFNKKLVISQISQI-PTTARMSYDNTGIRA-DVATDLFTADINHYTWSGDELMWL 139
QY 132 NSRSGVOPFGSQTATGVVAGHTWVWMOGQOTSMKIIISYVLTGPATISINLKAIFA 191
Db 140 ARYGVQVOPIGSQIAT-ATVQOQWELWYGANGSKYTSFVAPRTYTFQOQ-DVNDPFXYL 197
QY 192 AA-RGSLNTSDYLLD-----EAGEIIMOGGGLGNSFSV 226
Db 198 TQNHGFPASSQYLTITQFTEPTGPTATLSVNSMASVQOAGFEPMONGAGLAVNSFSS 257
QY 227 SV 228
Db 258 TV 259

RESULT 8
US-09-216-295-16
; Sequence 16, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Wendt, Dan J.

QY 183 DLKAFD--AARGSLNTSDYLDVAGEFEIMOGGGLGNSRFSVS 228
Db 214 NIRD-FEYNIADSKOMLSKTKYVSVFEGTEGTCQINISNDYV 260

RESULT 11

US-08-032-848C-12
Sequence 12, Application US/08032848C
Patent No. 5475101
GENERAL INFORMATION:
APPLICANT: Ward, Michael
APPLICANT: Clarkson, Kathleen A.
APPLICANT: Weis, Geoffrey L.
APPLICANT: Larens, Edward
APPLICANT: Lorch, Jeffrey D.
TITLE OF INVENTION: Purification and Molecular Cloning of
TITLE OF INVENTION: EG III Cellulase
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 180 Kimball Way
CITY: South San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/032,848C
FILING DATE: MAR 17 1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Horn, Margaret A.
REGISTRATION NUMBER: 33,401
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415 742-7356
TELEFAX: 415 742-7217
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 233 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-032-848C-12

Query Match 17.2%; Score 214.5; DB 1; Length 233;
Best Local Similarity 30.3%; Pred. No. 1.5e-12;
Matches 67; Conservative 33; Mismatches 90; Indels 31; Gaps 11;

QY 16 DGEYVOTNEMNSA---OQCLTINTATGA---WTVSTANFSGGTGAPATYPSLYK 66
Db 14 NKKYYLFNNWKGDEIKMGQOTIFYNSPISGMNMHMPSTHSYK-----AYPSLYS 65
QY 67 GCHNGNCTTKNVGMPIDISGSAVTSMTQVSSGAYDAVDYDWTNSTPTTQGPNGT- 125
Db 66 GWHHTACYTERSGPLRIDLSKRSITSVYTIKATGYTMAAYDTPHTTDDKANDSSPTD 125
QY 126 EIMIMLSRGVOPFGSQATGTVVAGHTNWNVOG-----QOTSMTIISVLTTPGATSI 179
Db 126 EIMIMLNDT-NAGPAGYIEI-VFLGDSMNVEFKGWTNNADNGGGMVFSFVHSGTNSA 183
QY 180 SNLDLKAIFADAAARGLNTSD--YLLDVEAGEFEIMOGGCG 218
Db 184 S-LNIRH-FTDYLVOTKQMSDEKISSVERGEIEF--GGDG 221

RESULT 12
US-09-216-295-20

Sequence 20, Application US/09216295
Patent No. 6268328
GENERAL INFORMATION:
APPLICANT: Mitchelson, Colin
APPLICANT: Wendt, Dan J.
TITLE OF INVENTION: No. 6268328e1 Variant BcIII-Like Cellulase Compositions
FILE REFERENCE: GC555
CURRENT APPLICATION NUMBER: US/09/216,295
CURRENT FILING DATE: 1998-12-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 20
LENGTH: 246
TYPE: PR1
ORGANISM: Emericella deeseroru
US-09-216-295-20

Query Match 16.8%; Score 208.5; DB 4; Length 246;
Best Local Similarity 31.3%; Pred. No. 5.8e-12;
Matches 71; Conservative 41; Mismatches 80; Indels 35; Gaps 12;

QY 17 GEYVOTNEM---NSSAOCLTINTATG---AMVYSTANFSGGTGAPATYPSLYKCHW 70
Db 41 GNFIVYNNLMGQDNADSGOTGVDSANGNSISMH-TTWSMGSGLSSVYS-----YANAY 94
QY 71 GNCCTKNVGMPIQISGSAVTSW---STQVSSGAYDAVDYDWTNSTPTTQGPNGTE 126
Db 95 QFTSTK-----LNSLSITFSMKWQSTTDIYA---NVAIDFTSS---SAGDSREYE 141
QY 127 IMIMLSRGVOPF---GSQATAGTVVAGHTNWNVOGQOTSMTIISVLTTPGATSI 183
Db 142 IMIMLAIGAGAPISSTGSIAT-VTLGVTWMSLYSGPNSGMQYSSFVASTTEESF-AD 199
QY 184 LKAIFADAAARGLNTSDYLDVAGEFEIMOGGCG-LGNSRFSVS 229
Db 200 LMDFTNYLAENQGLSSQYLTHVQAGTEPFTGATTLVSSVSYSVS 246

RESULT 13

US-09-216-295-12
Sequence 12, Application US/09216295
Patent No. 6268328
GENERAL INFORMATION:
APPLICANT: Mitchelson, Colin
TITLE OF INVENTION: No. 6268328e1 Variant BcIII-Like Cellulase Compositions
FILE REFERENCE: GC555
CURRENT APPLICATION NUMBER: US/09/216,295
CURRENT FILING DATE: 1998-12-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 12
LENGTH: 238
TYPE: PR1
ORGANISM: Fusarium equiseti
US-09-216-295-12

Query Match 16.6%; Score 206.5; DB 4; Length 238;
Best Local Similarity 31.7%; Pred. No. 8.5e-12;
Matches 71; Conservative 28; Mismatches 86; Indels 39; Gaps 14;

QY 8 ONGVTSVDDGEYVOTNEM---NSSAOCLTINTATGA-WTVSTANFSGGTGAPATY 61
Db 23 QYGLIS-SDG-YSLNNVNWKGDSGTGDCCTHVNNANAGACMDVEM-NMSGKDNV-RSY 78
QY 62 PSITKCHWGMCTKNVGMPIQISGSAVTSMTQVSSGAYDAVDYDWTNSTPTTQGPNGTE 113
Db 79 PN-----SALLIGEDKKTISITRMQSTAEKXSGDWLRDVAVDLEFTA 122
QY 114 STPTTQGPNGTEIMIMLSRGVOPFGSQATGTVVAGHTNWNVOGQOTSMTIISVLT 173

Db 123 ADPNHETSSEYELMWLARIGVOPIS-LQTSVTEGHTWELWGMNGSMKVFSEV-A 180
Qy 174 BEATISINDLKAIFADAARSLNTSD--YLLDVEAGFEIMOG 215
Db 181 PPPVNNFNADIKQ-FWDYLTCKSONEPADNOYLLTFQFGTEPPTG 223

RESULT 14
US-09-216-295-24
; Sequence 24, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; TITLE OF INVENTION: No. 6268328el Variant EgIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Erwinia carotovora
US-09-216-295-24

Query Match 16.4%; Score 204.5; DB 4; Length 263;
Best Local Similarity 30.5%; Pred. No. 1.5e-11;
Matches 67; Conservative 33; Mismatches 89; Indels 31; Gaps 12;

Qy 16 DGEYVQTEWMSA---QOCLINTATGA-----WTYSTANFSGTGGAATYPSIYK 66
Db 46 NKKYLLFNWVGKDEIKWQOTIFYNSPISMGMMHWPSTHSVK-----AYPSLVS 97
Qy 67 GCHMGCTKKNVMPIOISQISAVTSMSTQVSSGAYDAVDIMTNSPTTTGQPNGT- 125
Db 98 GWHWTAGTYENGLPIQLSSNKSITSNVTYSIKATCTYNAADIMPHITDKANWDSPTD 157
Qy 126 ETIMWLSRSGVQPPGSGATGTVVAGHTVWVWOG-----QOTSMTKIIISYVLTGATGIS 180
Db 158 ELMIMWINDT-NAGPAGDYLET-VFLGDSMNVEFKWINADNGGWN-VSFVHTSGTNSAS 214
Qy 181 NDLKAIFADAARSLNTSD--YLLDVEAGFEIMOGSOG 218
Db 215 -LNIIRH-FTDYLVQTKWMSDEKYSVEFGEIR-GDGG 251

RESULT 15
US-09-216-295-7
; Sequence 7, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; TITLE OF INVENTION: No. 6268328el Variant EgIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aspergillus kawachii (2)
US-09-216-295-7

Query Match 16.4%; Score 204; DB 4; Length 239;
Best Local Similarity 32.9%; Pred. No. 1.5e-11;
Matches 73; Conservative 26; Mismatches 99; Indels 24; Gaps 11;
Qy 17 GEYRVQTEWMS--SSAQOCLINTATGAMTV-STA-NFSGGTGGAATYPSIYKGCING 71

Db 31 GYTTINNMLMGDAGSGSCCTTVNSASAGTSMSTWMSGGENSVKS-----YAN 81
Qy 72 NCTTKNVGMPIOISQISAVT-SMSTQVSSGAYDAVDIMTNSPTTTGQPNGT 130
Db 82 SGLSFNKKLVSDISHIPTAARMSYDNTCIRRG- --AYDLFTAADINHTWSGDYELMIM 138
Qy 131 LNSRGVOPFGSQTATGTVVAGHTVWVWOGQOTSMKIIISYVLTGATGISINDLKAIFAD 190
Db 139 LARYGGVOPPLGSOIAT-ATVEGQTEWELWGVNGAQKTSFVAANPTTSPQG-DINDFEKY 196
Qy 191 AAA-RGSLNTSDYL--LDVEAGFEIMOGGOG--LGSNSFVSV 228
Db 197 LTONHGFPASSQYLITLALQFGTEPPTGGPATLNVADMSASV 238

Search completed: August 29, 2002, 16:22:38
Job time: 368 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:16:25 ; Search time 127 Seconds

(without alignments)
1074.005 Million cell updates/sec

Title: US-09-917-384-1

Perfect score: 6525
Sequence: 1 MERTQSGRCRCRYRGCTTRM.....GPNSTSGTNAAPTLCSTAS 1228Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq.032802:*

1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1023	15.7	1010	19	AAW34989 Tereidibacter end
2	1000	15.3	596	17	AAW34989 Tereidibacter end
3	999	15.3	423	17	AAW34989 Tereidibacter end
4	708.5	10.9	1751	20	AAW13493 Truncated cellul
5	668.5	10.2	381	21	AAW13493 Truncated cellul
6	668.5	10.2	381	21	AAW13493 Truncated cellul
7	663.5	10.2	1426	21	AAW13493 Truncated cellul
8	651.5	10.0	429	21	AAW13493 Truncated cellul
9	622.5	9.5	386	21	AAW13493 Truncated cellul
10	620	9.5	371	21	AAW13493 Truncated cellul
11	620	9.5	371	21	AAW13493 Truncated cellul

12	620	9.5	371	21	AAW13493 Truncated cellul
13	618	9.5	371	21	AAW13493 Truncated cellul
14	618	9.5	371	21	AAW13493 Truncated cellul
15	574.5	8.8	395	20	AAW13493 Truncated cellul
16	573	8.8	473	20	AAW13493 Truncated cellul
17	565.5	8.7	476	19	AAW13493 Truncated cellul
18	565.5	8.7	476	19	AAW13493 Truncated cellul
19	548	8.4	471	6	AAW13493 Truncated cellul
20	547	8.4	521	22	AAW13493 Truncated cellul
21	543	8.3	521	17	AAW13493 Truncated cellul
22	543	8.3	521	22	AAW13493 Truncated cellul
23	543	8.3	562	21	AAW13493 Truncated cellul
24	543	8.3	562	21	AAW13493 Truncated cellul
25	540	8.3	471	16	AAW13493 Truncated cellul
26	540	8.3	471	16	AAW13493 Truncated cellul
27	535	8.2	457	18	AAW13493 Truncated cellul
28	493.5	7.6	5179	22	AAW13493 Truncated cellul
29	488	7.5	1853	19	AAW13493 Truncated cellul
30	485.5	7.4	446	20	AAW13493 Truncated cellul
31	456	7.0	531	18	AAW13493 Truncated cellul
32	444.5	6.8	449	19	AAW13493 Truncated cellul
33	431.5	6.6	432	21	AAW13493 Truncated cellul
34	427.5	6.6	782	12	AAW13493 Truncated cellul
35	419	6.4	260	20	AAW13493 Truncated cellul
36	419	6.4	260	21	AAW13493 Truncated cellul
37	419	6.4	260	21	AAW13493 Truncated cellul
38	417.5	6.4	551	18	AAW13493 Truncated cellul
39	413.5	6.3	616	20	AAW13493 Truncated cellul
40	412	6.3	459	19	AAW13493 Truncated cellul
41	409	6.3	493	20	AAW13493 Truncated cellul
42	409	6.3	493	20	AAW13493 Truncated cellul
43	404	6.2	476	21	AAW13493 Truncated cellul
44	399.5	6.1	406	22	AAW13493 Truncated cellul
45	386	5.9	1352	22	AAW13493 Truncated cellul

ALIGNMENTS

RESULT 1	AAW34989 standard; Protein: 1010 AA.
XX	AAW34989;
XX	21-MAY-1998 (first entry)
DE	Tereidibacter endoglucanase.
XX	
KW	Endoglucanase; cellulase; carboxymethylcellulose; cellulose;
KW	biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;
KW	thermostable enzyme; thermophilic; glycosidase.
OS	Tereidibacter sp. (Clone 42GP1).
XX	
PN	WO9744361-A1.
XX	
PD	27-NOV-1997.
XX	
PF	22-MAY-1997; 97WO-US08793.
XX	
PR	22-MAY-1996; 96US-0651572.
PA	(RECO-) RECOMBINANT BIOCATALYSIS INC.
XX	
PI	Lam DE, Mathur EJ;
XX	
DR	WPI; 1998-018435/02.
DR	N-PSDB; AAT94197.
XX	
PT	Endoglucanase(s), preferably form archael bacterium, AEP11 1a -
PT	useful to degrade carboxymethylcellulose and hydrolyse of
PT	beta-1,4-glycosidic bonds in cellulose

XX Claim 1; Fig 1E; 164pp; English.

PS This protein comprises an endoglucanase of *Teredinibacter* (clone

CC 426P1) that is capable of degrading carboxymethylcellulose and of

CC hydrolysing the beta-1,4-glycosidic bonds in cellulose. It has

CC homology to an endoglucanase of archaeobacterium ABRP11a (see

CC AAM34985). It can be produced from native cells or from recombinant

CC host cells, especially prokaryotic host cells transformed with a

CC plasmid or virus-derived vector including the endoglucanase DNA

CC (see AAT94197). 24 Endoglucanases (see AAM34986-W35008) are claimed.

CC They can be used to degrade cellulose for the conversion of plant

CC biomass into fuels and chemicals, for use in detergents, textiles,

CC animal feed, waste treatment, and in the fruit juice and brewing

CC industries for the clarification and extraction of juices.

XX Sequence 1010 AA;

SQ

Query Match 15.7%; Score 1023; DB 19; Length 1010;

Best Local Similarity 48.5%; Pred. No. 6.5e-41;

Matches 210; Conservative 68; Mismatches 121; Indels 34; Gaps 13;

QY 57 VDNPPAGA-TFEVNPYMAEOVSEANOTNATLAAMRVSTYSTAVWMDRIATANGVNG 115

DB 598 vdnppagaqkyinpmws-----asaaneqgs-----vianeptfwmndrigalepad 647

QY 116 GPGITTYIDALISQOOGTTPREYIEIYIDLPGRCALALASNGELPATAGLIOTYETQYTD 175

DB 648 gmgldrhinealaq----gadlfmfvvyd]purdcaalasangelsedgfnlyksdyla 703

QY 116 PLASTLSPKXSLRIVYIIIEPDSLPMATNMSIQAC--ATNVPYEEGIEFALTKLNI 233

DB 704 plvelslpbyagiklaavlewdslpnlvtlsepdcqaangpgyrdgdirhatelgk1 763

QY 234 PNVTIYMDAASGWLGMPPNN-ASGYVOEQKYLNASIGVNGIDGVTYNTANTPKEPPM 292

DB 764 pnvytyvdahagvlgwmdnfagynlllyevvaangsglnplagtsansaytprpep1 823

QY 233 -TATQOVGGQPVESANFYQWNPDIADYAVDLYSRLVAAGPSSIGMLIDTLRNGWGP 351

DB 824 pdanfygvgqyrssdfyewasylakpfvldwrsamlskmpsslgmlidtarngwgp 883

QY 332 NEPTGPSTATDVTNVNOSKIDLRHGMCMONGAGLGQPPQASPTDPRNHLAYWI 411

DB 884 erptagssnmnltrvnesrltrhrgmwcnpq-gvgyrptlaa---pspgidayvw 938

QY 412 KPGESEDGTSAS--DPT-TGKKSDDPMCDP---TYTTSYGLTNALPNSPIAGOMFPA 463

DB 939 kpgesedgvsdpnfeldpndpukqhdpmcdpfasnsnsayg--tgampnphagrwfpe 996

QY 464 QPDQLVANNRPAY 476

DB 997 afglllenappt 1009

RESULT 2

AAR90715

ID AAR90715 standard; Protein; 596 AA.

XX AAR90715;

AC

XX 06-APR-1996 (first entry)

DT

XX

DE Thermostable cellulase-E3.

XX

XX Cellulase-E3; Thermomonospora fusca; thermostable; plasmid ps24;

KW plasmid ps26; *Escherichia coli*; cloning; *Streptomyces lividans*;

KW papain; catalytic domain; cellulose; chitosan; protease resistance;

KW synergism; cellolobiondrolase; beta-glucosidase; saccharification;

KW surfactant additive; paper recycling; delinking; paper refining.

XX

OS Thermomonospora fusca strain YX36.

XX Key Location/Qualifiers

EH Peptide 1..38

FT Peptide /note= "Signal peptide"

FT Peptide 39..48

FT Domain /note= "N-terminal peptide (AAR90717)"

FT Domain 174..596

XX /note= "Catalytically active domain (AAR90716)"

PN W09600281-A1.

PD 04-JAN-1996.

PP

XX 23-JUN-1995; 95MO-US09069.

PR 24-JUN-1994; 94US-0265429.

XX

PA (CORR) CORNELL RES FOUND INC.

PI Walker LP, Wilson DB, Zhang S;

XX

DR WPI; 1996-068865/07.

DR N-PSDB; AAT15595.

XX

PT Cellulase from *Thermomonospora fusca* - is thermostable and is useful

PT in a variety of industrial applications e.g. clarification of fruit

PT juices, fabric softening, etc

XX Claim 1; Page 32-37; 53pp; English.

CC The sequence represents thermostable cellulase-E3 (EC-3.2.1.4) from

CC *Thermomonospora fusca*, and is encoded by inserts in plasmid ps24

CC and plasmid ps26, obtained by cloning in *Escherichia coli*. The

CC gene may be cloned and expressed in *E. coli* or *Streptomyces*

CC *lividans* for recombinant cellulase-E3 production. Cellulase-E3,

CC or a catalytically active polypeptide obtained by papain digestion,

CC hydrolyses cellulose or chitosan at pH 5-11 and 40-70 deg C, and

CC has significant activity at over 60 deg C. E3 has higher stability

CC to proteolysis in culture supernatant than *T. fusca* cellulases-E2

CC and -E5, and shows strong synergistic activity when combined with

CC other cellulases, cellolobiondrolase and optionally beta-glucosidase.

CC The cellulase and mixtures may be used e.g. in cellulose

CC saccharification and ethanol production, fruit juice clarification,

CC as a surfactant additive, or in delinking or refining of recycled

CC paper.

XX

SQ Sequence 596 AA;

Query Match 15.3%; Score 1000; DB 17; Length 596;

Best Local Similarity 47.7%; Pred. No. 4.6e-40;

Matches 210; Conservative 55; Mismatches 143; Indels 32; Gaps 9;

QY 50 PAIAATHVDNPPAGATFEVNPYMAEOVSEANOTNATLAAMRVSTYSTAVWMDRI 109

DB 172 ptnpgekvdpiegaklynpwpsakaaeegs-----aaneastavwldrlga 221

QY 110 ING---VNGPGGLTYTYIDALISQOOGTTPREYIEIYIDLPGRCALALASNGELPATAAG 165

DB 222 iegndspitgsmglldhleav-rqsggplrlqyvllylpprdcaalasangelpdpde-- 278

QY 166 LOTYETQYIDPLASTLSN-PKYSRLRIVYIIIEPDSLPMATNMSIQACATAVPY----- 218

DB 279 ldrtyseyldpladlmwdfadyenrlrivaltsidspnlvtnvvgngygtelcaymknqng 338

QY 219 YEOGIEYALTKLHAIPNVIYMDAASGWLGMPPNNASGYVOEQKYLNAS-IGVNCIDOF 277

DB 339 yngvgyalrklqgelpnvyvlydaahgwljwdsnsgpsvdlfiyeaanngstlvdvngf 398

QY 278 VNTNATVYTLKPPFMATQOVGGQPVESANFYQWNPDIADYAVDLYSRLVAAGPSSIG 337

DB 399 lantanysatvepyldvngvtngqllrsgskwvdmngyvelsfvqdlrallakfrsdi 458

Oy		338	GMLIDTTLNMGCGNEPPESTADIVMTFVWQSKIDLRHGHGLMCNONGAGIGCPPOASP	397
Dd		459	gmldttenrgsggnprptpgssscldntlyvesrldrrlhpncwcnqagaglgscrclvp	518
Oy		398	TDFPNALDAIVTWIKPGESDGTSAASDPPTGKKSDPMCEPTYTSYGVLTN--ALPNS	454
Dd		519	a----pygdvayvwvkppgesdgaseelndegkxgfdrmcdptygnarnngnpsgalpna	574
Oy		455	PIAGOWFAOEDOLVANARP	474
Dd		575	pigswhsaqdtrellanayp	594
 RESULT 3 AAR90716 standard; Protein: 423 AA. AAR90716 AAR90716: AAAR90716: 06-APR-1996 (first entry) Thermostable cellulase-E3 catalytic domain.				
XX			Cellulase-E3; Thermomonospora fusca; thermostable; papain;	
KW			catalytic domain; Escherichia coli; cloning; Streptomyces lividans;	
KM			cellulose; chitosan; protease resistance; synergism;	
KM			cellobiohydrolase; beta-glucosidase; saccharification;	
KX			surfactant additive; paper recycling; delinking; paper refining.	
XX			Thermomonospora fusca strain YK36.	
XX			WO9600281-A1.	
PN			04-JAN-1996.	
PD			23-JUN-1995; 95MO-US09069.	
Pf			24-JUN-1994; 94US-0265429.	
PR			(CORR) CORNELL RES FOUND INC.	
PA			Walker LP, Wilson DB, Zhang S;	
XX			WPI; 1996-068865/07.	
DR			N-PADB; AAT15596.	
PT			Cellulase from Thermomonospora fusca - is thermostable and is useful	
Pr			in a variety of industrial applications e.g. clarification of fruit	
PP			juices, fabric softening, etc	
PS			Claim 1; Page 38-40; 53pp; English.	
XX			The sequence corresponds to a catalytically active domain of	
CC			thermostable cellulase-E3 (EC-3.2.1.4) from Thermomonospora	
CC			fusca, and is isolated by papain cleavage of the full-length	
CC			protein (AAR90715). The binding domain and linker region have	
CC			been removed from the full-length protein. The protein may be	
CC			produced in recombinant form in Escherichia coli or Streptomyces	
CC			lividans. The domain retains full activity, but shows reduced	
CC			cellulose binding. The catalytic domain, or the corresponding	
CC			full-length cellulase, hydrolyses cellulose or chitosan at pH	
CC			5-11 and 40-70 deg C, and has significant activity at over 60	
CC			deg C. E3 has higher stability to proteolysis in culture	
CC			supernatant than T. fusca cellulases-E2 and -E5, and shows strong	
CC			synergistic activity when combined with other cellulases,	
CC			cellobiohydrolase and optionally beta-glucosidase. The cellulase	
CC			and mixtures may be used e.g. in cellulose saccharification for	
CC			ethanol production, fruit juice clarification, as a surfactant	
CC			additive, or in delinking or refining of recycled paper.	
XX			Sequence 423 AA:	
XO				

[illegible]


```
XX DR WPI, 1999-395187/33.
XX PT EGIII like cellulase
XX PS
XX Example: Fig 6; 47pp; English.
CC The present polypeptide represents a full-length sequence of a
CC novel EGIII-like cellulase of Streptomyces lividans (Ce1b). It was
CC deduced from a gene sequence isolated from genomic DNA using PCR
CC primers (see AAY59180-91) based on conserved motifs (see AAY06325-29)
CC of Trichoderma reesei EGIII cellulase and related enzymes. PCR
CC has been used to identify novel EGIII-like enzymes, including the
CC present protein, from bacterial and fungal sources (see AAY06331-70).
CC The sequence shows homology to T. reesei EGIII (see AAY06330). Also
CC provided by the invention are vectors, host cells and methods
CC for the recombinant production of such enzymes, which can be used
CC in the treatment of cellulose-containing textiles, as feed
CC additives, in the treatment of wood pulp, in the reduction of
CC biomass to glucose, in the stone washing of indigo dyed denim, or
CC as laundry detergent components (all claimed).
XX SQ
XX Sequence 381 AA:
XX
XX Query Match 10.2%; Score 668.5; DB 20; Length 381;
XX Best Local Similarity 39.1%; Pred. No. 1.7e-24;
XX Matches 158; Conservative 49; Mismatches 150; Indels 47; Gaps 13;
XX
XX 829 AGAGTAPSPRYATTTSPSPPTPTGTTVDCTPGPNONGVTSVGDDEYRQTNEMNSS 888
XX 17 algaavlaafalvaslvtlaapagadt---fclcep---fgtltlqg-ryvvgnmrwgt 67
XX
XX 889 AAOCLTINTATGAMTVSTANFSGTGAPATYPSIKYCHMGNCTTKNVMPIQISIGS 948
XX 68 apqcvta-ldtg-fvrtqadgsaprtngapkysvfnghytrc-spgtldprldtvs 124
XX
XX 949 AVTSMSTTOYSSGAYDAVDIWTNSTPTTGGPNGTETIMIMLNSRGCVOPPGSQTATGVT 1008
XX 125 apaslsygfvdgavnyasylwldptarldg-vnqteimlwfnrvpqlqpsvgt-as 182
XX
XX 1009 VAGHTNNVWOGQOTSKWIIISYVLTGATISINLDKAIKIPDAARGLSNTSDVLDYEAG 1068
XX 183 vgrtvevwsygnsgndvlsfv-apsalsgwsldvmdfvratvarglaendwyltsvq 241
XX
XX 1069 FEIMOGGOGIGSNSFSVSTSGTSSPTSPSPPTPTSPSPPTSPSPSPSSSG 1128
XX 242 fepwngaglaavnsfssvcltvgpgt-----dpgdpggpps--- 277
XX
XX 1129 VACRATVYVNSDMGSGFTATVTNTGSRATNGMTVAMSPFGNQTYNNVNTALTOGAS 1188
XX 278 -acavsygtuv-wqdgftadvtnltgtpavdgwqlafllpsqgrltlnaanaaltpssgs 335
XX
XX 1189 VTATNLSYNNVIOPGOSTTFEGFNGSYSGTNAAPT-----LSCST 1226
XX 336 vfatgashnarlapygslsfqgltyggafaepclgfringlact 379
XX
XX RESULT 6
XX AAB14881
XX ID AAB14881 standard; Protein; 381 AA.
XX
XX AC AAB14881;
XX
XX DT 21-NOV-2000 (first entry)
XX
XX DE Rhodothermus marinus EGIII-like cellulase.
XX
XX KW Rhodothermus marinus; Trichoderma reesei; endoglucanase III; EGIII;
XX cellulase; mutant; enzyme stability; textile treatment;
XX wood pulp treatment; feed additive; detergent.
XX
XX OS Rhodothermus marinus.
```

```
XX PN WO200037614-A2.
XX PD 29-JUN-2000.
XX
XX PF 12-NOV-1999; 99MO-US26704.
XX
XX PR 18-DEC-1998; 98US-0216295.
XX
XX PA (GEMV ) GENENCOR INT INC.
XX
XX PI Mitchinson C, Wendt DJ;
XX
XX DR WPI; 2000-482483/42.
XX
XX PT Novel endoglucanase III or endoglucanase III-like cellulase useful for
XX treating textiles and wood pulp comprises a substitution or deletion at
XX specified positions in the wild form of endoglucanase III -
XX
XX PS Example 1; Fig 3; 52pp; English.
XX
XX CC The present sequence is a cellulase related to endoglucanase III (EGIII)
XX from Trichoderma reesei. EGIII-like genes were isolated from genomic DNA
XX libraries constructed from various microorganisms by PCR. The isolated
XX genes showed significant homology to EGIII from T. reesei. Certain
XX substitution and deletion mutations have been incorporated into EGIII and
XX EGIII-like cellulases to produce variant enzymes with improved stability,
XX e.g. increased resistance to temperature stress. The mutants may be used
XX in textile and wood pulp treatment, as a feed additive, and for reducing
XX biomass to glucose. They are also useful for stonewashing or indigo dyed
XX denim and as an agent in laundry and dish detergents.
XX
XX SQ
XX Sequence 381 AA:
XX
XX Query Match 10.2%; Score 668.5; DB 21; Length 381;
XX Best Local Similarity 39.1%; Pred. No. 1.7e-24;
XX Matches 158; Conservative 49; Mismatches 150; Indels 47; Gaps 13;
XX
XX 829 AGAGTAPSPRYATTTSPSPPTPTGTTVDCTPGPNONGVTSVGDDEYRQTNEMNSS 888
XX 17 algaavlaafalvaslvtlaapagadt---fclcep---fgtltlqg-ryvvgnmrwgt 67
XX
XX 889 AAOCLTINTATGAMTVSTANFSGTGAPATYPSIKYCHMGNCTTKNVMPIQISIGS 948
XX 68 apqcvta-ldtg-fvrtqadgsaprtngapkysvfnghytrc-spgtldprldtvs 124
XX
XX 949 AVTSMSTTOYSSGAYDAVDIWTNSTPTTGGPNGTETIMIMLNSRGCVOPPGSQTATGVT 1008
XX 125 apaslsygfvdgavnyasylwldptarldg-vnqteimlwfnrvpqlqpsvgt-as 182
XX
XX 1009 VAGHTNNVWOGQOTSKWIIISYVLTGATISINLDKAIKIPDAARGLSNTSDVLDYEAG 1068
XX 183 vgrtvevwsygnsgndvlsfv-apsalsgwsldvmdfvratvarglaendwyltsvq 241
XX
XX 1069 FEIMOGGOGIGSNSFSVSTSGTSSPTSPSPPTPTSPSPPTSPSPSPSSSG 1128
XX 242 fepwngaglaavnsfssvcltvgpgt-----dpgdpggpps--- 277
XX
XX 1129 VACRATVYVNSDMGSGFTATVTNTGSRATNGMTVAMSPFGNQTYNNVNTALTOGAS 1188
XX 278 -acavsygtuv-wqdgftadvtnltgtpavdgwqlafllpsqgrltlnaanaaltpssgs 335
XX
XX 1189 VTATNLSYNNVIOPGOSTTFEGFNGSYSGTNAAPT-----LSCST 1226
XX 336 vfatgashnarlapygslsfqgltyggafaepclgfringlact 379
XX
XX RESULT 7
XX AAY13492
XX ID AAY13492 standard; Protein; 1426 AA.
XX
XX AC AAY13492;
```

[illegible]

OY		786	NVGAHVNVVRNGVLVGQPTVTSTDTDGLAGATAYTYTAAVDAAAGNTABSPPTYATT	845
Dd		594	- - - - -cprrcat- - - - -adlptvrsalp	611
OY		846	SPSPSPPTGTGVTVDCRPG- - - - -PNQN- - - - -GWTSVGGDEX	878
Dd		612	tpraptaspygssy--wpssesyalakvwyanagnlsaptnlvnpkikienvgttav--dlS	667
OY		879	RVCQTNEW---NSSAQCLLTINTATAGANTYSTANFSGGTGAPRTYSIK--GCHWNC	932
Dd		668	rkvkrwywylidgeatgsvsassinpalyld-vklganagadyvelgtfksagaylaaq	726
OY		933	TTKNVMPIQLSQGSASVTSMSITTOVSSGAVDYADAYDIWNISPTTTG-	979
Dd		727	sckelrfsj- - - - -qkgsgsyngndy--svrsalcylenekvtglyldd	768
OY		980	- - - - -OPN-GTEIMIW- - - - -INSRGVQPFQSCATATGVTAAGHTMNMWOGQ	1020
Dd		769	vlywgrepstnaqlkwvyangnlspntvnlpnkikienngtctavdsrvkvrywtlidge	828
OY		1021	QTSWKIISTYLTPCATSISNLDKAIFADAARSLNTSDYLLDV--EAGFEIIWGQG--	1076
Dd		829	atgsvsfcsinsp----ayldvkfvklganagg--ddyvelgtfksagaylaagst	879
OY		1077	- - - - -GIG- - - - -SNSFSV- - - - -SVT- - - - -SGTSSP	1094
Dd		860	keltslsdqkgsysyndySlrsansyenekvtglyldgalwmgrepsrgtkpkagvvlr	939
OY		1095	TPSPSPPTPSPTPTPPSPSPSPSP- - - - -SEPTSSP- - - - -SSSGACRAFTYYVNSDDMGSGFTA	1147
Dd		940	tpaptcstpcpplttccptcptvcvrplsccptpvssclpctacbpctcpsilclpap	999
OY		1148	TYVYTNTGSRATNGMTVAAMSFEG-GNOTVNYMWNTALTQSASVYATATNISNNVIOPGST	1206
Dd		1000	tatpctrpps-vtddnddwlfaggnkiv-----dkdgkpwxlltgvmw-----	1040
OY		1207	TFGFNGSYSGTN 1218 :	
Dd		1041	fghn-tgcn 1048.	
RESULT 8				
ID	AAV84346			
XX	AAV84346 standard; Protein; 429 AA.			
AC				
XX	AAV84346;			
DF				
XX	12-JUL-2000 (first entry)			
DE	Amino acid sequence of an endoglucanase III (Egitl)-like cellulase.			
XX				
KW	Endoglucanase III; Egitl; Egitl-like cellulase; surfactant stability;			
KW	cellulase; textile processing; textile cleaning; stonewashing;			
KW	indigo dyed denim; cellulose containing fabric; fabric smoothness;			
KW	pill removal; fibril removal; cotton; cellulostic fibre; dying; detergent			
KW	animal feed; wood pulp; paper; grain; biomass reduction; glucose.			
OS	Streptomycetes lividans.			
XX				
PN	WO200014208-A1.			
XX				
PD	16-MAR-2000.			
XX				
PZ	24-AUG-1999; 99WO-US19154.			
PR	03-SEP-1998; 98US-0146729.			
XX				
PA	(GENEV) GENENCOR INT INC.			
XX				
FI	Fowler T;			
DR	WPJ: 2000-271052/23.			

XX Novel variant endoglucanase III-like cellulases with improved
PT surfactant stability and resistance to temperature stress, useful for
PT textile processing or cleaning, treating wood pulp, food and grain, and
PT reducing biomass to glucose
XX
XX
PS Disclosure: Page 65-66; 73pp; English.
XX
XX The present sequence represents an endoglucanase III (EGIII)-like
CC cellulase. The cellulase has homology to the Trichoderma reesei EGIII
CC protein. The variant cellulases have improved temperature stability,
CC and improved surfactant stability. The variant cellulases and
CC compositions containing them are used in textile processing or cleaning,
CC e.g. stonewashing of indigo dyed denim, and modifying the texture, feel
CC or appearance of cellulose containing fabrics (e.g. improving fabric
CC smoothness or removing pills and fibrils). The compositions may also be
CC used for the removal of immature or dead cotton from cellulosic fibres
CC or fabric, which can cause uneven dyeing. The cellulase may also be used
CC in a detergent composition for washing laundry and dishes and in the
CC treatment of animal feed, wood pulp, paper, non-animal foods and grains.
CC The enzymes may also be used in the reduction of biomass to glucose.
XX
XX
SO Sequence 429 AA:

Query Match 10.0%; Score 651.5; DB 21; Length 429;
Best Local Similarity 36.9%; Pred. No. 1.2e-23;
Matches 165; Conservative 48; Mismatches 169; Indels 65; Gaps 15;

QY 813 GLAAG-----TATYTVAAVDAAGNTSAPST---PYTATTT-----TSPS-P-849
DB 13 gllaalgavlaafvslvsaapagadtlcpefgltlqgrvynrwgslapqcv 72
QY 850 SPTPTGTTTDCPTGPNONGV-----TSVGDDEXRYOTNEMNSAQOCLTINTATGAWTS 905
DB 73 laldqftrtqdgdsaprtgarkyspvtlncgylqnmwgsclapcvta-ltlg-fryc-130
QY 906 TANFSGGTGAPATYPSIYKGMHNGCTTKNVMPIQISGASVSWSTTQVSSGAYDV 965
DB 131 qdgdsaprtgarkyspvtlncgylcnc-spgtdlpyrltlvtasapslsygfvdgavyna 189
QY 966 AYDIWTNSTPTTGGOPRTEIMTLNRSKGVOPFGSQTATGTVACHTWNVMWOGQOTSMK 1025
DB 190 bydlwldprlartldg-vnqkeltmlwlnrvyrlqplspvgtl-asvgrtlwvsgngsnd 247
QY 1026 IISYVITPATSISNLDKAIFADAARGLNLSYDLIDVEAGFEIMOGGSGNSFSV 1085
DB 248 vslstv-apalslsgwslvmdlfratvargl-aendwyltavgagfepwqngaglaavnsfs 306
QY 1086 SVTSGTSSPTSPSPPTPTSPSPPTSPSPSPSSSGVACRATYVNVSDMGSGF 1145
DB 307 tvetlsgpgtlt-----dpbgdpgps-----acavsygtlrv-wqdgf 340
QY 1146 TATYTVTNTGSRATNGVWAMSEFGNQTYTNTWNTALTOSGASVATNLSTYNNVIOPGOS 1205
DB 341 ladvltlntgltltpvdgwqlaflpsqgtiltnawnaaltpssgvtatgshnarlapgs 400
QY 1206 TTEGFGNSYSGTNAAPT-----LSCT 1226
DB 401 lsfqfgytgygaafepctgflrlnctact 427

RESULT 9
AAV67497
ID AAV67497 standard; Protein; 386 AA.
XX AAV67497;
XX
XX 19-MAY-2000 (first entry)
XX
XX Problem sequence of Cella and cellulase 11A68 fusion.
XX
XX Cellulase; Actinomycetes; detergent; feed additive; textile treatment;
KW

KW pulp; paper; cellulase 11A68.
XX
XX Streptomycetes lividans.
OS Synthetic.
XX
XX
FH Key Location/Qualifiers
FT Peptide 1..46
FT Protein /note="Cela signal sequence"
FT 47..386
FT /note="cellulase 11A68 mature peptide"

WO200009707-A1.

24-FEB-2000.

28-MAY-1999; 99WO-US11971.

24-JUN-1998; 98US-0104308.

18-NOV-1998; 98WO-US24649.

28-MAY-1999; 99US-0321981.

(GEMV) GENENCOR INT INC.

Jones BE, Van Der Kleijf WAH, Van Solingen P, Weyler W;
WPI; 2000-224344/19.

N-PsDB; AA57031.

A novel Actinomycetes cellulase and related DNA, useful for detergent
compositions, treating cellulases and paper or pulp -
Example 6; Fig 15; 72pp; English.

The invention provides a cellulase from Actinomycetes. The cellulase can
be used in a detergent composition, as an additive for animal feed and
for the treatment of textiles or pulp and paper. The DNA encoding the
cellulase can be used to identify homologous cellulases and for
recombinant production of cellulases. The present sequence represents
the protein sequence of a cela signal sequence and cellulase 11A68
fusion sequence contained in the expression cassette consisting of the
GI promoter, cela signal sequence, cellulase 11A68 and GI terminator
XX
XX
SO Sequence 386 AA:

Query Match 9.5%; Score 622.5; DB 21; Length 386;
Best Local Similarity 38.6%; Pred. No. 2.7e-22;
Matches 152; Conservative 51; Mismatches 132; Indels 59; Gaps 18;

QY 854 TGTVTIDCT--PQP-----GVTSGDDEXRYOTNEMNSAQOCLTINTATGAW 902
DB 29 tgyvslvgltalalappagangqldrygttlqg-dryvqnmrwgtsatgc--lnvngnft 85
QY 903 TVSTANFSGGTGAPATYPSIYKGMHNGCTTKNVMPIQISGASVSWSTTQVSSGA 962
DB 86 eltqdgsvtlnqgarkyspvtlncgylcncaprtl-lpmrlslsgaapsvayrfyng 144
QY 963 YDVAAYDIWTNSTPTTGGOPRTEIMTLNRSKGVOPFGSQTATGTVACHTWNVMWOGQOT 1022
DB 145 ynaaydwlwldprlmg-varletmlwlnrvpvpqplspvgtl-ahvgrfswewtqng 202
QY 1023 SMKIIISYVLTGATSIISNLDKAIFADAARGLNLSYDLIDVEAGFEIMOGGSGNS 1081
DB 203 sndvlsf-lapsalswsvfvdg-fvdqavshglatpdylltslqngfepwgggtglavn 260
QY 1082 SFSVSVTS-GTSSPTSPSPPTPTSPSPPTSPSPSSSGVACRATYVNVSD 1140
DB 261 sfssavnaagsgngtlp-----tpa-----acqvsyelt- 280
QY 1141 WSGGFTATYTVTNTGSRATNGVWAMSEFGNQTYTNTWNTALTOSGASVATNLSTY-NV 1199
DB 291 wpgsfvtdtlntgltltpvdgweldflpagnltvtaamallspagavlatarsatgngrl 350

DT 08-AUG-2000 (first entry)
XX Amino acid sequence of a C1 endoglucanase polypeptide.
DE
XX
XX C1 endoglucanase; promoter; mutant; expression-regulating region;
KW secretion signal sequence; enzyme; cellulase; xylanase; pectinase;
KW lipase; protease.
XX
OS Chrysosporium lucknowense.
FH
FT Key Location/Qualifiers
FT Peptide 1..28 /note="signal peptide"
XX
XX W0200020555-A2.
XX
XX 13-APR-2000.
XX
XX 06-OCT-1999; 99MO-NL00618.
XX
XX 06-OCT-1998; 98MO-EP06496.
XX
XX (AARL-) AARL INC.
XX
XX Emailarb MA, Burlingame RP, Olson PT, Sinitzyn AP, Parriche M;
XX Bousson JC, Pynnonen CM, Punt PJ, Van zellj CMJ;
XX
XX WPI: 2000-317705/27.
XX N-PSDB: AAA14818.
DR
PS Disclosure; Page 43-46; 92pp; English.
XX
XX The present sequence represents a Chrysosporium C1 endoglucanase. The
CC polynucleotide sequence comprises a promoter sequence, and can be
CC linked to a nucleic acid sequence of the invention to control
CC expression. The specification describes a mutant Chrysosporium
CC strain which comprises a nucleic acid sequence encoding a polypeptide
CC of interest, where the nucleic acid sequence is linked to an
CC expression-regulating region and a secretion signal sequence.
CC The mutant strain is useful for producing heterologous proteins
CC and polypeptides of interest, e.g. industrially useful enzymes such
CC as cellulases, xylanases, pectinases, lipases and proteases.
XX
SQ Sequence 395 AA:

Query Match 8.8%; Score 574.5; DB 21; Length 395;
Best local Similarity 34.7%; Pred. No. 5.3e-20;
Matches 160; Conservative 58; Mismatches 156; Indels 87; Gaps 16;
QY 15 RGTTPMAISKRLRAGVLAGVSNASIVPLAMQHAIATHVNDNRYAGATFFVNPYMAQ 74
DB 21 rttcpqkp---rqsag--casavtlas-----tnvfqy---tlhpnnyra 60
QY 75 EVOSEAAQNTNATLAAMRYVSTYAVWMDRIAINGVNGPEGLTTYLDAAISQOOGTT 134
DB 61 eveaaaeaisdsaleakrvadvgtflwidtlenlg-----rlepal--edvpc 108
QY 135 PEVIEIIVYDLPGRDCAALASNGELPATAAGLQTFEROYIDPITASILSNPKYSLRIVTI 194
DB 109 enivgavivypidpdcakasangel--kygelidtykcyldkilaellk--ahsntafalv 164
QY 195 IEPDLSLPNATNNSIQACATAVPYEEGIEYALTKLHAIPNVIYMDAHSGLWGPENNA 254
DB 165 ltpdsalpnltvnsdlqtcqgsaagyregyalqnl-lpnyvmyldagbhgwlgwdanl 223
QY 255 SGYVOEQAKLVNASIGVNGIDGFTVTNTANTTPLEKPEFMTATQOVGGQPVESANFYQWNP 314
DB 224 kpsagelasyvksagspgyrsltnvagw-----nawdqepe--fadaadaqynkc 274

QY 315 IDEADYAVDLYSRLVAAFPSSIGMLIDTLRNMGWGPNEPTGPSTADVNTFVNQSKIDL 374
DB 275 qneklyintfgaelksagmpnh--alldtgrngvtg-----1 309
QY 375 RQHRGLMCNONGAGLGOPPOASPTDFPNAILDAYVWIKPPEESDGTSAASDPTTGKSDP 434
DB 310 rdewgdwcwnvgagfgvyrplantgd---eladatvwkpggsdgtadsa----aaryds 362
QY 435 MCDPTTTSYGVLTNALPNSPIAGWFPAPQFDLVANARPA 475
DB 363 fccgk-----pdatkpspeagtlwnqayfemllknanps 394

Search completed: August 29, 2002, 16:21:40
Job time: 315 sec

...

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:18:25 ; Search time 72.78 seconds
(without alignments)
1621.293 Million cell updates/sec

Title: US-09-917-384-1

Perfect score: 6525

Sequence: 1 MERTQOSGRNCRYOGRITRM.....GFNGSYSGTNAPTLSTAS 1228

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 100%

Listing first 45 summaries

Database : PIR_71:*

1: pirl:*

2: pirl:*

3: pirl:*

4: pirl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1655.5	25.4	872	2 A49541	cellulase - Cellul
2	1358	20.8	683	2 A82704	1,4-beta-cellulobios
3	1051	16.1	579	2 T35240	probable secreted
4	1000	15.3	596	2 A55976	cellulase 1,4-beta
5	744.5	11.4	1742	2 T17120	cellulase (EC 3.2.
6	743.5	11.4	1711	2 T31337	1,4-beta-glucanase
7	675	10.3	1779	2 T31085	xyylanase - Caldice
8	671.5	10.3	1331	2 A48954	mannan endo-1,4-be
9	649.5	10.0	382	2 JC2571	cellulase (EC 3.2.
10	641	9.8	438	1 S70602	cellulose 1,4-beta
11	581.5	8.9	1090	2 S59077	cellulase 1,4-beta
12	580	8.9	1045	2 A39199	endoglucanase B (E
13	550	8.4	2232	2 T34434	hypothetical prote
14	548	8.4	471	1 A26160	cellulose 1,4-beta
15	548	8.4	471	1 A38979	cellulose 1,4-beta
16	519.5	8.0	1039	2 S02711	cellulase (EC 3.2.
17	505	7.7	747	2 B47093	cellulase (EC 3.2.
18	488	7.5	1854	2 S36859	c1pa protein - Clo
19	473.5	7.3	473	2 S50755	hypothetical prote
20	471	7.2	3020	2 A43932	mucin 2 precursor,
21	443	6.8	1367	1 S48478	glucan 1,4-alpha-g
22	416.5	6.4	474	2 S15921	protein TPX-VT3 -
23	411	6.3	279	2 T10361	hypothetical prote
24	405.5	6.2	915	2 A43802	cellulase (EC 3.2.
25	404.5	6.2	1275	2 T33369	hypothetical prote
26	404	6.2	441	2 T12011	cellulase (EC 3.2.
27	398.5	6.1	1032	2 T34433	hypothetical prote
28	393.5	6.0	4776	2 E95206	cell wall surface
29	387.5	5.9	1585	2 T31611	hypothetical prote

30	386	5.9	145	2 A41897	cellulase homolog
31	385.5	5.9	1777	2 T34369	hypothetical prote
32	382.5	5.9	3570	2 T45025	mucin MUC5B, trach
33	379	5.8	1459	2 T32271	hypothetical prote
34	378	5.8	508	2 G69593	cellulase (EC 3.2.
35	377.5	5.8	700	2 B41897	cellulase (EC 3.2.
36	375.5	5.8	499	2 JN0111	cellulase (EC 3.2.
37	375	5.7	1483	2 C97012	probably celluloso
38	374.5	5.7	351	2 S50754	hypothetical prote
39	370.5	5.7	508	2 A26874	cellulase (EC 3.2.
40	369.5	5.7	1664	2 T18262	S-layer protein
41	369	5.7	890	2 T35237	probable secreted
42	366	5.6	499	2 A27198	cellulase (EC 3.2.
43	363.5	5.6	2271	2 F90073	hypothetical prote
44	359	5.5	1104	2 S59310	probable membrane
45	355.5	5.4	380	2 G70847	probable cellulase

ALIGNMENTS

RESULT 1

S49541

cellulase - Cellulomonas flm1

C:Species: Cellulomonas flm1

C:Date: 29-Nov-1995 #sequence-revision 01-Mar-1996 #text-change 22-Oct-1999

C:Accession: S49541; A47093

R:Meinke, A.; Gilkes, N.R.; Kwon, E.; Kilburn, D.G.; Warren, R.A.J.; Miller Jr., R.C.

Mol. Microbiol. 12, 413-422, 1994

A:Title: Cellulohydrolase A (Cbha) from the cellulolytic bacterium Cellulomonas flm1

A:Reference number: S49541; MUID:94344030

A:Accession: S49541

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-872 <ME1>

A:Cross-references: EMBL:L25809; NID:9456028; PID:NAC36898.1; PID:9456029

R:Meinke, A.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.

J. Bacteriol. 175, 1910-1918, 1993

A:Title: Cellulose-binding polypeptides from Cellulomonas flm1: endoglucanase D (Cend

F:769-870/Domain: bacterial cellulose-binding domain homology <BCB>

F:770-869/Disulfide bonds: #status predicted

A:Accession: A47093

A>Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 812-872 <ME2>

A:Experimental source: ATCC 484

C:Superfamily: bacterial cellulose-binding domain homology

F:769-870/Domain: bacterial cellulose-binding domain homology <BCB>

F:770-869/Disulfide bonds: #status predicted

Query Match 25.4%; Score 1655.5; DB 2; Length 872;

Best Local Similarity 33.1%; Pred. No. 2,4e-62;

Matches 408; Conservative 125; Mismatches 300; Indels 399; Gaps 28;

```
Db 249 TLFAEYAK--STTAGFASIDGFSVDVANTTPLEEPLSDSSSLTINNTPIRSKFEYEMNF 306
Qy 315 IDEADYAVDLXRLVAGPSSSGMIDLRNGMGPNPTGPSTADVTNTPVNSKIDL 374
Db 307 FEIDIDTAHMRLLVAAAGPSSIGMLVDSRNGMGPNPTSTATAVDNAYADARVGR 366
Qy 375 ROHRGLMONGAGLGPQASPTDEPNNAHLDAYVWIKRPGESDGTSAASDPPTGKSP 434
Db 367 RYHRGAMCPPLAGLGRFEPATPSGAASHLDAFWIKRPGESDGTIDPDQGRFR 426
Qy 433 KCDPVTYTS--YGVLTNALPNSPIAGQWPAQFDOLVANARAPVTSTSSSPSPSPS 492
Db 427 KCDPFTFVSKLNNQLGAPPNAPLAGQWFEQFVTLVKNAPVIGGT-----FVEDL 479
Qy 493 ASPSPSPSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSS 552
Db 480 VAPT----- 483
Qy 553 PSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 612
Db 484 -----VPTGLTAGT 492
Qy 613 TSSSSVDLSTVTRKWTFRDGSSTLVYNCNMAAGCGRIRASFGSVNPAFTADTYL 672
Db 493 TTATSVPLS-----WTA-----STDNAAVTGYDYVRCGT 520
Qy 673 SFTGGLAGSGTGEIONVNSKSNFTETNDYSGTNTTFQDMKVVYVNGVYVNGT 732
Db 521 TLVGTAA-----TSYTVTGLTPATV--SFTVRK 549
Qy 733 EPSGSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 789
Db 550 DAAGVNSAASAAATGSGVLTDTAPSVPAGLTAGTITTTTTPSLMVASIDNAGSGV 609
Qy 790 AATYTRKRVLVGPVTSFTDTGLAGTAITYTAAVDAAGTASAPSPVAT----- 843
Db 610 AGYEVLKGTVVGTATVTSYVGLTAGTTSFSYAKVAGTSAASAASATQGTGV 669
Qy 844 --TSPS--PSPPTGTTVDCTPGPNQNCVTSGQDEYVQINEMNSAQCCLITATAG 900
Db 670 VTTTASVPTGLTAGTITTSVP-----LT----- 694
Qy 901 AATVSTANPSGSGTGAATPSIYKCHWNGCTKNVGMPIQISQISAVTSMTQVSS 960
Db 695 --WTASIDN--AGSGVA----- 708
Qy 961 GAYDAVYDWTSTPTTGTGPNGTETIMILNSRGGVQPFSGQATGVVAGHTWVWQ 1020
Db 709 -----GYEFGNGTTRVAT-----VTSSTYTVTG----- 731
Qy 1021 QTSWKIISVLTGATISINLKAIFADAARGLSNTSDYLLDVNAGFEIMQGGGLS 1080
Db 732 -----LAADTAISFTYKADVAGVSA----- 754
Qy 1081 NSFVSVTSIGTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSS 1140
Db 755 ---SAAVASRTOAAT-----SG--CCTVYKASS- 778
Qy 1141 WSGGTAITYTNTGSRATNGTVAASFGCNOTVTNTNTALTQSGASTATINLSYNNV 1200
Db 779 NWTGFGVGVVNNGTAAALNGWTLGFSFADGQKVSQSGMSAEMWSQSTAVTAKNAPNGT 838
Qy 1201 OPGOSTTFEGNGSYSTNAAPT-----LSC 1226
Db 839 AAGSSVTSIGFNGTHTNATPAFTLNGVACT 870

RESULT 2
AB2704
1.4-beta-cellobiosidase XF1267 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: AB2704
```

```
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: AB2515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: AB2704
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-683 <SIM>
A:Cross-References: GB:AE003960; GB:AE003849; NID:g9106242; PIDN:AAF84076.1; GSPDB:GN
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Artuda, P.; Abreu, F.A.; Ascencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carrero, D.M.; Carre
as Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Klajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, R.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, C.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovsky-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1267

Query Match 20.8%; Score 1358; DB 2; Length 683;
Best local Similarity 43.9%; Pred. No. 5.2e-50;
Matches 319; Conservative 81; Mismatches 218; Indels 108; Gaps 19;

Qy 53 AATHVDNYPAGATFEVNYPAQDEVOSEAANQNTATLAARKRVYSTAVMDRIAING 112
Db 18 AEAHVDNPEVDATSYLNPDIYSKIDITSIKAVNDVSLKEKQITKNPTGVMTIDTAVAG 77
Qy 113 VNGGPG---LFTYTLDAALSOQCGTPEVLEIYIYDLPGDCAAALNSGLPATAGLQTY 169
Db 78 GPRNPGRLNLTGHLDAALQKGNPTITATFYIYDIPGDBCHALNSGELPLPLEGLQRY 137
Qy 170 EQOYIDPIASILSNPKYSSRLIYTIIEPDLPAVYNNMSIOACATA--VPYEOGLEYVL 227
Db 138 KREYIDTASISNPKYIKRIYNTIIEPDLPIITNMSTPSCAQKKNIGYIEGKIKAL 197
Qy 228 TKLHAIPNVIYIWDAAHSGMLGPNNASGYVOEVQVL--NASIGVNGIDGYFNTANTYP 286
Db 198 NKLSEIPNVIYNYWDIGHSGMLGMDTRIPAVSLYTKVIGSTAGFASVNGPATNTANTYP 257
Qy 287 LKEPFM--TATQOYCGQPVASANTYOWNPDIIDEADYAVDLXRLVAAAGPSSIGMLIDTIR 345
Db 258 LIEPNLPNDLNTGQPIRSKFEYENRYFDEMDVSELYNDFVAAGWPSISIGFIIDTGR 317
Qy 346 NGMGEPNEPTGPSTADVTNTPVNSKIDLRHRLMNCNONGAGLGPQASPTDPNNAHL 405
Db 318 NGMGEPNPTS--AFGNDVASYNSGRIDRRHNGMNCNKDAGIGLPIAT-----PGHY 372
Qy 406 DAYVWIKRPGESDGTSAASDPPTGKSPDCPTTYTTSYGVLTNALPNSPIAGQWPAQF 465
Db 373 DAFOWIKRPGYSSSLIPNDQKGFDRYCDPTFTTPGVLTGALPDLGDMWHAQF 432
Qy 466 DQLVANARAPVPTSTSSPPPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 525
Db 433 VKLNNAYPDI-----SKTPPLGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 486
Qy 526 SPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 585
Db 487 SGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 546
Qy 586 GGLKLYOKNNDAPGDNQIKPGLQLVNTGSSVDLSTVTRKWTFRDGSSTLVYNCDMA 645
Db 547 SG-----SSSSSSSSSS-----SSGSSS-----GSSSGSGFN----- 572
```



```

QY      646  AMGCC-NIIRAFSGVSNPATPTADTLYQLSFTGCTIACAGSGEIQNRNKSQDMSNFTETN 704
Db      573  -PGAGFNPGAGF---NP-----GAGFNPGAGSDPTKNPIT-----SSITVDS 610
QY      705  DYSVG-----TNTTFQDPTKVTYVNGVL--VMGT----- 732
Db      611  DMHTGYCGRVAKVTNIGSSRSSMT-VTIPLKG7IQIPLMSATWSLSSGDXKLNASGLDMNKTLL 669
QY      733  EPSTGS 738
Db      670  EPNGTPT 675

RESULT      3
T35240
Probable secreted cellulase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: J35240
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z21572
A:Accession: J35240
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-579 <Seq>
A:Cross-references: EMBL:AL031515; PIDN:CAA20645.1; GSPDB:GN00070; SCOEDB:SC5C7.33
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC5C7.33

```

Query Match	16.1%	Score 1051:	DB 2:	Length 579:
Best Local Similarity	49.2%	Pred. No. 3.1e-37:		
Matches 215:	Conservative	61:	Mismatches 131:	Indels 30: Gaps

QY	50	PAIAATHVNDNPYAGATFFVNPYMAOEVOSEANOTNATLAAKMRVSTYSTAVMDRRIA	109
Db <td>158</td> <td>PTGPEDRVDNPNFDGAQVYVYNPEMSANAAAEPEGDR-----IADEPTGVMDRRIA</td> <td>207</td>	158	PTGPEDRVDNPNFDGAQVYVYNPEMSANAAAEPEGDR-----IADEPTGVMDRRIA	207
QY <td>110</td> <td>INGVVGSGGLTTLDDAALSQOOGTTPREYIEYITDLPGRDCAALASNGELPRTAAGLOTY</td> <td>169</td>	110	INGVVGSGGLTTLDDAALSQOOGTTPREYIEYITDLPGRDCAALASNGELPRTAAGLOTY	169
Db <td>208</td> <td>IEGANGGGLRDHIDDAAL-EOKGSEEMVQVITVYMLPGDCSALASNGELPTE--IDRY</td> <td>264</td>	208	IEGANGGGLRDHIDDAAL-EOKGSEEMVQVITVYMLPGDCSALASNGELPTE--IDRY	264
QY <td>170</td> <td>ETQYIDPIASISLNPKSYSLRTVETIIEPDSLNPATVNNMSIQACAT-----AVPYEEG</td> <td>222</td>	170	ETQYIDPIASISLNPKSYSLRTVETIIEPDSLNPATVNNMSIQACAT-----AVPYEEG	222
Db <td>265</td> <td>KTEYIDPIAEILISDSKYADRLRTVETIELDSLPNLTNYSGRFATIEDNCVMKANGYQAG</td> <td>324</td>	265	KTEYIDPIAEILISDSKYADRLRTVETIELDSLPNLTNYSGRFATIEDNCVMKANGYQAG	324
QY <td>223</td> <td>IEVALTKLHAIPNVYIYMDAAHSGMLGMPNNAAGYVOEKVLMASIG--VNGIDGFTVN</td> <td>280</td>	223	IEVALTKLHAIPNVYIYMDAAHSGMLGMPNNAAGYVOEKVLMASIG--VNGIDGFTVN	280
Db <td>325</td> <td>VGYALNRLGAVGNVYNYVDAGHHGMLGMSDN-FGPSAEIEFKTAATTEGATLDDVHGFTVN</td> <td>383</td>	325	VGYALNRLGAVGNVYNYVDAGHHGMLGMSDN-FGPSAEIEFKTAATTEGATLDDVHGFTVN	383
QY <td>281</td> <td>TANTYPLKEPFPFATTAQQVGGGQRPESANFQWMPDIDDEADYAVDLSRLVAAGFPSISGL</td> <td>340</td>	281	TANTYPLKEPFPFATTAQQVGGGQRPESANFQWMPDIDDEADYAVDLSRLVAAGFPSISGL	340
Db <td>384</td> <td>TANYSALKEENFKITDYSVNGTSVROSDDWDMQYTDIELSYAAMRKLVLSLDFDNLGML</td> <td>443</td>	384	TANYSALKEENFKITDYSVNGTSVROSDDWDMQYTDIELSYAAMRKLVLSLDFDNLGML	443
QY <td>341</td> <td>IDTLNKGNGGPRPEPSTATDVNTFVNSOK-IDLROHRLKCNONGAGLGOPQASPTDF</td> <td>400</td>	341	IDTLNKGNGGPRPEPSTATDVNTFVNSOK-IDLROHRLKCNONGAGLGOPQASPTDF	400
Db <td>444</td> <td>IDTISNKGNGGADRPPGPGATITDVNTFVNGRGRDKRKHILGKMCNOSAGIGERPQASPA--</td> <td>501</td>	444	IDTISNKGNGGADRPPGPGATITDVNTFVNGRGRDKRKHILGKMCNOSAGIGERPQASPA--	501
QY <td>401</td> <td>PNALHDAVYVIMKPGESDGTSAASDPTTGKSKSDPKCDPYT--TSYGLVTNALPNSPA</td> <td>457</td>	401	PNALHDAVYVIMKPGESDGTSAASDPTTGKSKSDPKCDPYT--TSYGLVTNALPNSPA	457
Db <td>502</td> <td>--AGIDAVYVIMKPPGESDSSKLTINDPKDGFDRKCDPYTGTNERNGNSMGSLPAPIS</td> <td>559</td>	502	--AGIDAVYVIMKPPGESDSSKLTINDPKDGFDRKCDPYTGTNERNGNSMGSLPAPIS	559
QY <td>458</td> <td>GGMFPAGQFDQVLAANRP</td> <td>474</td>	458	GGMFPAGQFDQVLAANRP	474
Db <td>560</td> <td>GAMFSAQFOELMKKNAYP</td> <td>576</td>	560	GAMFSAQFOELMKKNAYP	576

RESULT	4
A55976	
cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) - Thermomonospora fusca	
C:\Species: Thermomonospora fusca	

```

C:Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 22-Oct-1999
C:Accession: A55976
R:Zhang, S.; Lao, G.; Wilson, D.B.
Biochemistry 34, 3386-3395, 1995
A:Title: Characterization of a Thermomonospora fusca exocellulase.
A:Reference number: A55976; MUID:95186496
A:Accession: A55976
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-596 <ZHA>
A:Cross-references: GB:18978; NID:9664822; PIDN:AAA62211.1; PID:9664823
A:Superfamily: bacterial cellulose-binding domain homology
A:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:40-140/Domain: bacterial cellulose-binding domain homology <BCB>
F:41-139/Disulfide bonds: #status predicted

Query Match      15.3%; Score 1000; DB 2; Length 596;
Best Local Similarity 47.7%; Pred. No. 4,4e-35;
Matches 210; Conservative 55; Mismatches 143; Indels 32; Gaps

```

	Query Match	15.3%	Score 1000:	DB 2:	Length 596:	
	Best Local Similarity	47.7%:	Pred No.4,4e-35:			
	Matches 210;	Conservative 55;	Mismatches 143;	Indels 32;	Gaps 9	
OY	50 PAIAATHVDNPYAGATFEVNPYMAOEVOSEANQTNATLAAKRRVSTYSTAVWMDRIA	109				
	: : : : : :		:			
Db	172 PNPPEKEDVNPEEGAKLYXNPNWSAKAAAEPCGS-----AVANESTAWIDRIGA	221				
OY	110 ING-----VNGSGGLTYYIDAALSGOOGTFREVEIETIYIDLPGRDCAALSNEELPATAG	165				
	: : : :					
Db	222 IEGNDSPPTGSKGLDHDIEEAV-KOSGGDLPTIQVIYTWLPGBDCALASNSGLGPDE--	278				
OY	166 LQTYETQYIDIPIASILSN-PKYSLSRIYVIIIEPDSLPPNAVTMMSIOACATAVPY-----	218				
Db	279 LDRTSEXYIDPLADIMMDEPADYEENIRIVAIIEIDSLPLNTLVNGNGTELCAYMKRONCG	338				
OY	219 YEOGLEVALTKHAIIPNYUYIMDAASHGSLGCPNNASGYVOEVOKYLAHS-IGVNIIDOF	277				
Db	339 YVNGVGIALRKIGEIPNYNYITIDAHHHGIGMDSNGSPSDIFFEYEAANSGSTVDDVHPF	398				
OY	278 VTNFNANTYPLKEPFMTAQOVGOGPYESANFQWNPDIDEADYADVLSRLVAAGFPSSI	337				
	: : : : : : : : : : : : : : : : :					
Db	399 ISNTNASTYATPELDVNGTVNGQLIRGSKVMDMNQYDELSEFYODLRQALIAKGFRSDI	458				
OY	338 GMLIDTLRNGMGCPMEPCPSFATQYNFMFVNOSKIDLQRHLKMNQONAGIGOPPOASP	397				
Db	459 GMLIDTSRNGMGCPNMPGPSSTIDNTANTYVDSRIDRKRHNPGWCMOAAGJGERITYNP	518				
OY	388 TDFPNAHLDAYWIKRPGEISDGTSAASDPDTTGKSDPMCDPTYTTSYGLVTN---ALPNS	454				
	: : : :					
Db	519 A----PGVDAYVWKVRPGESDGASEEIPNDECKGFBRMCDPTYOGNARGNPNPSGALLPWA	574				
OY	455 PIAGOWFPAQFDOLVANARP	474				
Db	575 PISGHMFSAQFRELLANAP	594				

```

RESULT      5
T17120
cellulase (EC 3.2.1.-) precursor, thermoactive - Caldocellum saccharolyticum
C:Species: Caldocellum saccharolyticum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C:Accession: T17120; A43745
R:Te'o, V.S.; Saul, D.J.; Bergquist, P.L.
Appl. Microbiol. Biotechnol. 43, 291-296, 1995
A:Title: Cella, another gene coding for a multidomain cellulase from the extreme therm
A:Reference number: Z18568; MUID:95356703
A:Accession: T17120
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1742 <TEO>
A:Cross-references: EMBL:L337742; NID:9537499; PID:9537500; PIDN:AA91086.1
R:Luethli, E.; Bhana Jasmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene cod
A:Reference number: A43745; MUID:91247819

```

A:Accession: A43745
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1516-1544, 'A', 1546-1742 <LUE>
A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72860.1; PID:g144293
C:Genetics:
A:Gene: cels
C:Keywords: glycosidase; hydrolase

Query Match 11.4%; Score 744.5; DB 2; Length 1742;
Best Local Similarity 22.8%; Pred. No. 7.5e-24;
Matches 281; Conservative 162; Mismatches 364; Indels 423; Gaps 46;

29 AGVLAGANSTASIVPLAMOHPAIATPV-----DNPYAGATFEVNPY- 71
173 AVVETAASILAASIVLDRNPTKATYLOAHKADLYEEFAEYTKSDSGYTAANGYVMSWG 232
72 -----MAQVQSEANONTNNTLAAMKRVSTSTAVVMMDRIAANGVNGRGGLTYIDA 125
233 FYDELSTMA-----AVWL--YLATND-----STYLTK 256
126 ALSQOQGTTPREIYIYDLPGRCALASNGE--LPATAAGLQTYETQYIDPIASILSN 183
257 AESYVQ-NMPKISGNIIDYKRAHCMDVHNGAALLAKITDKQYK-QIIESHLDYMTT 314
184 P-----KYSLRIYTTIEPDSLPAVNTMSI-----QACATAV-----PYEGQIEY 225
315 GYNGERIKYTPKGLAMLDOMGSLRYATTATLAFLAYSDMSGCPGCKEYRKFGESQIDY 374
226 AL-----TKLHAIPIVNYIYMDAHSGW-----LGMNNAASGYVQ 259
375 ALGSGRSEFVVGFGTNPKRPHHRTAHSWSADSOISFTHHRTLLYGALVGGSDSTYD 434
260 EVQKVLNASIGVNGIDGFTVTNTANY-----TPLEPPM----- 292
435 DISNVNNEVACDYAGVYALAKKYLXYGNNPIPFKAIEPTNDEFEVEAGINASGTN 494
293 -----TAQVQGGQGVESANFYQWNPDIDEADYANDLYSRLYAAGFPSSIGMLITLNG 347
495 FLEIKAIYNNOSGMPARATNKKFR-----YFVDL--SELKKAGY----- 532
348 WGGPEPTGPTATDVNFVNOSKIDLRQHRGLMCNONGAGLGGPQASPTDFPNAHLDA 407
533 ---SPNQLT-----LSNTNYD-----GAKYSGP----- 552
408 YVW-----IKPPGESDGTSAADPTTGKKSDDPMCDPYTTSYGVLTNAL 451
553 YWDSRNIYIYLVDFGTGLYPGQDKYKKEVGRIRAPQVQWMDNSNDYSFQDIKGV 612
452 PNPPLAGQWFAQDFQL-VANARPAVPTSTSSPPPPSPSASPSPSPSSSPSS 510
613 SGSVKTKTYILYDEDDIVMGEDEP-----TSGVSTP 644
511 PPSPPSPSPSPSPSSPSSPSPSPSPSPSPSSPSSPSSPSPSPSPSPSPSPSPSP 570
645 PTAATPTPTPT-----TPATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 688
571 PSSSPSPSPSPSP--VSGGLKVOYKNNDSAPGDNQIKPGLQLVNTSSSSVDLSTVYVRV 629
689 PTVPT 748
630 TRDQSSSLVYNCORAMAGCGNIRASFGSVNPAVPTADTYLQLSTGCT--LAAGSGTGE 687
749 TYDGERAASALS-DWAQIGASNVTFKFWKLSSVSAGADYLEIGFKSAGQLQPKDKE 807
688 IONRNKSKDMSNFEETNTSY-GLNTTFQDMTKYVVYVNGVYVWGTESGTSPTSPS 746
808 IOIRNKDMSYNGNDMSWYIQSMTSYGENEKVAYIDGVLVWQDEPSGTPAPPTSTPT 867
747 PPSPPSPGCVTPSPVPTGLVYTVGVSGSSVSLAMNASTDNVGAHNYRNGVLGQOPTV 806
868 VT-----VTPTPTPT-----PTV 880

807 TSFTDTGLAGTAYTYVAAVDAAGNTSAPSTPVATTTSPSPSPPTPGTVDCTPGPN 866
881 T-----PPTVAT--PPTPTPTPTSPS----- 902
867 QNGVTSVQDEFRV-----QTNENSSAOQCL-TINTATGANTVSTANFSGGTGAPATY 920
903 ---TPATGGQIKVLANKETINSTNTIRPWLKVNSSGSSIDLRSVTR----- 948
921 PSIVKCHWGNCTYRNVGPIQISGAVTSMSTVOSSGADVAAYDWTNSTPTTQ 980
949 -----YW-----YTVDGERAQSALSDMA-----QIGASNTYFK-FVRLSSVSQA 987
981 PNGTITMTWLSRGVQDPGSCQATGVTYVAGHTV-NWQGGQTSKILISYLLTPGATSIS 1039
988 DYELBIG-FKSGAGQLQPKDKEGIQIRPNKDMENYNGNDMSW-----IQSMTSYG 1039
1040 NIDLKATFPDAARSLNTSDVLLDVEAGFEFWOGGGLGSSFSVSYSSTSPSPS 1099
1040 ENEKVATAYD-----GVLVW-GQEPGAPAPVPTPTPTPTPTPTPTPTPTPTPTPT 1080
1100 PPT 1127
1081 PPT 1110

RESULT 6

1.4-Beta-glucanase (EC 3.2.1.-) - Anaerocellum thermophilum (fragment)
C:Species: Anaerocellum thermophilum
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T31337
R:Zverlov, V.; Maht, S.; Riedel, K.; Bronnenmeier, K.
Microbiology 144, 457-465, 1998
A:Title: Properties and gene structure of a bifunctional cellulolytic enzyme (CelsA).

A:Reference number: 221003; MUID:98154434
A:Accession: T31337
A:Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1711 <ZVE>
A:Cross-references: EMBL:286105; NID:el071329; PID:es350354; PIDN:CAB06786.1
C:Genetics:
A:Gene: celsA
C:Keywords: glycosidase; hydrolase

Query Match 11.4%; Score 743.5; DB 2; Length 1711;
Best Local Similarity 22.8%; Pred. No. 8.1e-24;
Matches 277; Conservative 169; Mismatches 353; Indels 417; Gaps 47;

35 AVSIASTVPLAMOHPAIATVNDNPAAGATFVNPVMAOEVSBA-ANQTNATLAAMKR 93
156 AASLAASIVLDRNPTKATYLOH-----AKELYEFAEYTKSDSGYTAANG 202
94 VVSTVS-----TAVWMDRIAANGVNGRGGLTYIDALASQOQGTTPREIYIYDL 145
203 YNWSNGFYDELSTMAAVWL--YLATND-----STYLTKAESYVQ-NMPKISGNTIDY 252
146 PRCDCALASNGE--LPATAAGLQTYETQYIDPIAS-----ILSNP-----Y 186
253 KRAHCMDVHNGAALLLAITGDKIYK-QIIESHLDYWTGYNGERIKYTPKGLAMLDOW 311
187 SSLRIYV-----ILEPDLPAVNTNMSIOACATAVPYEGQIEVAL-----TKLHA 232
312 GSLRATITATLAFLAYSDVGCPTSTKEIYR-----KRESIDYALSGAGSPVYVGET 366
233 IDNVYIYMDAHSGW-----LGMNNAASGYVQEVOKVLNASIGVNGI 274
367 NPPKRPHERHTAHSWSADSOISFTHHRTLLYGALVGGSDSYTDDISNVNNEVACDYN 426
275 DGFVNTANTYTPLEKPEPMATQVGGQPVESANFYQWNPDIDEADYANDLYSRLYAAGFP 334

Db 427 AGFVGALAK-----MYOLYGNPDPFKAIE--TPTNDEF-----FEVAGIN 466
 Qy 335 SS-----ICMLIDTLRNWGGNEPTGATDVNTFVNQSKIDLRQ--HRGLMCNQ-- 384
 Db 467 ASGTNFEIETAIYNN-OSGM-----PAKADKLKF--RYFDLSBELIKAGISPNQLT 515
 Qy 385 -----NGAGLGOPPOASPTDFPNAHLDAVW-----IKPGESEDGTS 422
 Db 516 LSTNYNAGAKVSGP-----YVMDASKNIYYILVDFGTGLIYPGGODKYK 560
 Qy 423 ASDPTCKKDDPMCDPTTYSYCVLTNALPNSPIAGQWFA--QPDQVLAARAPVPTSTS 481
 Db 561 EYGFRIAPONVQWMDNSNDYSFQDIKGVSSGVKRYIPLDYDDVWMDGPGTSGAT- 619
 Qy 482 SSPPPSPSSASPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPS 541
 Db 620 -----PPTATATPT 660
 Qy 542 SSPSPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSS 600
 Db 661 TPTPTATPTSTP-----TPTSTPSSTPVAGQIKVLYANKETNST 700
 Qy 601 DNOIKPELOLVNCGSSVDLSTVYVRYWFTRDGSSSTLYVNCMAAGCCGIRASFGSVN 660
 Db 701 TMTIRPMLKLVNCGSSSIDLSRVIRYVYDGDKAQSAIS--DMAQIGASVYTFKFKLS 759
 Qy 661 PATPTADTYLQLSFTGCT--LAAGSGTGEIQNRVKNKSDMSNFTETNDYSY--GINTTFQDM 717
 Db 760 SSVSGADYIEIGFKSAGQLOAKQKQDIQIRNKSNDMSYNGNDMSWMSQMTNGEN 819
 Qy 718 TKVTVYVNGVLWGTETSGTSPSPSPSPSPSPSPSPSGDVTTPSVPTGLVTVGVSSSV 777
 Db 820 VKVYAYIDGVLVWGOEFGSGATPTPTATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 855
 Qy 778 LANNAISDNVGVANHYNRNGVLVGOPTVTSFTDTGLAGATYTYVAADAAGNTSAPS 837
 Db 856 -----TPTPT-----S 861
 Qy 838 TPVATATTSPPSPPTPTPTPTDCTPGPNONGVTSVOGDEYV-----QTNEMNSSAQOC 892
 Db 862 TPTATPTATPTPTPTPTSTVA-----GGQIKVLYANKETNSTTIRPW 906
 Qy 893 L-TINTATGAMVYSTANFSGGTGAPATYPSIYKGMCTTKNVGMPLOISIGSAVT 951
 Db 907 LKVVNCGSSSIDLSRVIR-----YW-----YVADGDKAQAIS 940
 Qy 952 SMS-----TQVSGAYDAYDIWNTSTPTTGOPNGMEIM-----WLN 991
 Db 941 DMAQIGASNTYFKFKLSSSVSGADYILEIFKSGAGOLQAKDTG--EIQIRENKSQMSN 999
 Qy 992 SRGVOFGSGOTATGVTVAGHTWNVMOGQOTSW-----KIISYVLTPTGATISINDLKAIF 1047
 Db 1000 YN-----QGDMS--WMSMTNNGENKVTAYI----- 1025
 Qy 1048 ADAARGLSTSDYLDYERAGFEIWOGGGLGNSFSVTSSTGSPSPSPSPSPSPSPSP 1107
 Db 1026 -----GQVLVW--GOEFGSATPTPTATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 1065
 Qy 1108 PTPSPSPSPSPSPSS 1123
 Db 1066 PTAATPTPTPTPTPTPT 1081

RESULT 7
 T31085
 xylanase - Caldiceilusirupator sp.
 C:Species: Caldiceilusirupator sp.
 C:Date: 02-Sep-2000 #sequence_rev1sion 02-Sep-2000 #text_change 02-Sep-2000
 C:Accession: T31085
 R:Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.
 submitted to the EMBL Data Library, December 1997
 A:Description: Family 10 and 11 xylanase genes from Caldiceilusirupator sp. Rt69B.1.
 A:Reference number: Z20972

A:Accession: T31085
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1779 <MOR>
 A:Cross-references: EMBL:AF036924; NID:q2760905; PID:q2760909; PIDN:AA895326.1
 C:Genetics:
 A>Note: xynC

Query Match 10.38; Score 675; DB 2; Length 1779;
 Best Local Similarity 23.68; Pred. No. 6.3e-21;
 Matches 257; Conservative 150; Mismatches 416; Indels 266; Gaps 40;

Qy 169 YETQY-----IDPIASILSNPKYSLRIVYIIIEPDSLPNATNMISIQACATAVP-- 217
 Db 659 YKSKYATWMSLEPIVYVYVNSLTPAP-----AIQVPTSTPTPTPTPTPTPTPTPTPTPTPT 713
 Qy 218 -----YEOGIEYALTKLAIIPNVYIYMDAHSGLGWPNNASGYOEVOKVLNASTGV 271
 Db 714 SPAGSGYWTPESEYGAALKV-----WY-----ANGMSSTYTNVLPKIKI 752
 Qy 272 NGIDGFVNTNANTYPLKEPEMTATQVGGQPVESANFQWNP-----DIDBAD 319
 Db 753 ENVG--TTAVDSLVRKVRWYTTIDGEAQSVAAS--SIPAYIDVRVYKLGANAGGAD 807
 Qy 320 YAVDYSRLVYAGFPSSIGMLIDTLRNWGGNEPTGATDVNTFVNQSKIDLRQHRG 379
 Db 808 YVEV-----GFKSGAGV-----AAQST-----KEIRLSIQKSSG 839
 Qy 380 LMCNONGAGLQOPQASPTDFPNAHL--DAYWIKPGESE-----GTSASDPTGKKS 432
 Db 840 SYNOSNDYVSANSYIENEKVTGIDYLVWGEPEGHNAQIKWYANGNLSPTNVLNP 899
 Qy 433 DPMCDPTTTS-----YGLVFNALPNSPIAGQWFAQPD-----OLVANA----- 472
 Db 900 KIKIENGTIAVDLSRVKVRWYTTIDGEATQSVASASINAYIDVRVYKLGANAGGADY 959
 Qy 473 -----RPVPTSTSSPPPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 496
 Db 960 YVEYGFKSGAGVLAAGGSKTEIRLSIQKSSGSYNOSNDYSVANSYIENEKVTGYIDV 1019
 Qy 497 -----PSP 550
 Db 1020 LVWGEKPSRGTKPAGVETPAPT 1077
 Qy 551 SSP 609
 Db 1078 ATPPTPTPT--TPT 1135
 Qy 610 LVNCGSSVDLSTVYVRYWFTRDGSSSTLYVNCMAAGCCGIRASFGSVNATPTADTY 669
 Db 1136 IVNCGSSVDLSRVKIRWYVYDGDKAQSAIS--DMAQIGASVYTFKFKLSVSGADY 1194
 Qy 670 LQLSFTGCT--LAAGSGTGEIQNRVKNKSDMSNFTETNDYSY--GINTTFQDMTKVTVYVNG 726
 Db 1195 LEVGFSSGAGLOQKDTGDIQVRFNKNKDMGNYNQADMSWMSQMTNNGENAKVLYLYD 1254
 Qy 727 VLVWGTETSGTSPSPSPSPSPSPSPSPSGDVTTPSVPTGLVTVGVSSGSVLANNAISDN 786
 Db 1255 VLVWGOEFGATPAPTSTATPT 1314
 Qy 787 VGVANHYNRNGVLV-----GQPTVTSFTDTGLAGTAYTY-----TVAAVDAAGNTSAPS- 837
 Db 1315 VGLIIPN--NNPLISKHFGADPAVLVY-----GGRVYVNLTDILEYDENGAVKXMSY 1365
 Qy 838 ---TPVATATTSPPSPPTPTPTDCTPGPNONGVTSVOGDEY--RVQTNEMNSSAQOC 892
 Db 1366 SKINKIRIYSSDDLVNMTDHEIE--VAGP--NGVAKMANFSAVAVAYKIKINGKDRFF 1420
 Qy 893 LTINTATGAMVYSTANFSGGTGAPATYPSI-----YKGMH----- 929
 Db 1421 LYFNGSGGIGVITADSFPGWSDPLGRPLTWSTPGVGVVWLPDPAVLVYDDGKAYTY 1480

```

OY 930 -----GNCITKNYGMPIQ-----ISOIGSAVT-----SMSTOVSSGADVADYDITMTN 972
Db 1481 FGGVPGQDAMPNTARVLPQDDMI SYVGSAAITPAPYMEDSGINIGMTYYTSCYN 1540
OY 973 STPTTGQPNGETEIMWILNSRGVQFPGSQATGVY-----VAGHT----- 1013
Db 1541 FAQRGSGPAPGALIVMT-----GKSPVGPWEYRGVILNPGNFVGVGNHHLQLEFNCK 1596
OY 1014 WNWQGOQTSMKIISVLTPEGATS-----ISNIDKAIKPADAAKRSLSLTSYLLDVE 1066
Db 1597 WYIYVHAQTLKKDLG--VAKGYSRPHINQVOIENGIVIKETVADYKGAQVKKFDPRMVE 1654
OY 1067 AGEIWMGG 1075
Db 1655 AETPAMCAG 1663

RESULT 8
A48954
manan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum
N:Alternate names: beta-mannanase
C:Species: Caldocellum saccharolyticum
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: A48954; B43745
R:Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.
A:Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multidomain
A:Reference number: A48954; MUID:93119139
A:Accession: A48954
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1331 <GB>
A:Cross-references: GB:L01257; NID:g144290; PIDN:AA71887.1; PID:g144291
A>Note: sequence extracted from NCBI backbone (NCBIN:121576, NCBI:P121577)
R:Luehth, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding
A:Reference number: A43745; MUID:91247819
A:Accession: B43745
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337, 'PROHQRHQR' <LUE>
A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AA72861.1; PID:g144294
A>Note: the authors translated the codon CAC for residue 262 as Glu
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 10.3%; Score 671.5; DB 2; Length 1331;
Best Local Similarity 23.5%; Pred. No. 6.3e-21;
Matches 243; Conservative 145; Mismatches 305; Indels 341; Gaps 43;

OY 240 MDAHSGMLGWPNN-----ASGY-----VOEVQKV--LMSIGVNGI----- 274
Db 67 IDTALRGIRSMQNSVRVVLNSNGYRWTKIPASEVANITLSLSLGRKALILEVHDTGVC 126
OY 275 -DGEFTNTAN-----YTPLK-----EPWTATQVGGQVSEANATQAMPDIDEADYVD 323
Db 127 ECGAACSLAQAAYWEKERSVLGNEDEYIT--NIGNEFYGNNNYQWNVDTKMAIKA-- 182
OY 324 IYSRLVAAGFPSSIGMLIDTLRNG--WGGPNEPTGPS-----TA 360
Db 183 -----LRDAQFKRTI--WDAPRWGQDWSMTMDNOSIHEADPLRLNLYVSLIMYGYNTTA 236
OY 361 TDVNTFVNOSKIDLRHQRGLMCNONGAGIGQPPQASPT---DFPNAHLDAYVWIKPPGSS 417
Db 237 SKVEYFI-KSEYD-----KGL-----PLVIGFGHGHMDG-----DPDE 270
OY 418 DGTSAASDPTTKKSDPMC-DPTTYSIGVLTNALPNSFIA-GQWPPAQFDOLVANARPA 475
Db 271 AIVRYAKQYKIDLFMSWMSGNSGYGLDMVNNMPPNNPTPGCWYKTN-----A 320
OY 476 VPTSTSSPPPPPSASPSPPSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSS 535

```

```

Db 321 IGT----- 331
OY 536 SSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 594
Db 332 TYPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 373
OY 595 NDSAPGDNQIKFGDLVNTGSSVDLSVTVRYWFTTRDGSSTLVYNCDDMAAGCGNIRA 654
Db 374 KETNSTTWTIRFWLKVYVNGSSSIDLSRWTIRYWTVVGERSAALS--DMAQIGASNTVF 432
OY 655 SFGSVNPAFTPTADTYLOLSTFGCT--LAAGSTGTLORVYVNSDMNSNTEFTDYSY-GTN 711
Db 433 KFKVLSSSVGADYILEGFKGAGOLQPKDTGELQMFNFNDMSNTNQGDMWSIQSM 492
OY 712 TTFQDMTKVTVVNGVLVWGTGTSPTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 771
Db 493 TSYGENEKYTAIDGVLVWQGPSCGATPAPATATPTPTPT-----VTP----- 536
OY 772 SGSSVSLAMNASTDNGVAHVYVNRGVLVGQPTVTSFTDTGLAAGIATVTVAAVDAAG 831
Db 537 -----TPVT----- 541
OY 832 NTSAPSTPTATTTSPSPSPPTGTVTDCPTGPNQNGVTSVQGEYRY-----QTNEMN 886
Db 542 ---PTPTVTAT---PTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 583
OY 887 SSAQOL-TINTATGAMTVSTANFSGGTGCGAPATYPSIYKGMCGNCTTKVGMPIQISQ 945
Db 584 NTIRPMLKVYVNGSSSIDLSRVTIR-----YV-----TYVDEGR 617
OY 946 TCSATVSTTGVSSGADVADYDITNTSTPTTGQPNGETEIMWILNSRGVQFPGSQAT 1005
Db 618 AQSASIDMA---QIGASVNTFK-EVKLSSSVSGADYILEIG-FKSGAGOLQPKGDTGERI 671
OY 1006 GTVYAGHTM-NVWQGOQTSMKIISVLTPEGATSISNIDKAIKPADAAKRSLSLTSYLLD 1064
Db 672 QIRFKSDMSNTNQGDMWS-----IOSMTSYGENEKYVAYID----- 709
OY 1065 VEAGEIWMGGCGGLNSPFSVSTGTSPTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1124
Db 710 ---GVLVW-GQPSGCTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 765
OY 1125 SSSGVACRATTVYVNSDMSGFATVTV-TNTGSRATNGTVMASGNGQCTVNTYNTALT 1183
Db 766 TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 821
OY 1184 QSGASVTATNTLSYN 1197
Db 822 LQGNRLITGYNMENN 835

RESULT 9
JC2571
cellulase (EC 3.2.1.4) precursor - Streptomyces rochei (strain A2)
N:Alternate names: endo-1,4-beta-glucanase; endoglucanase
C:Species: Streptomyces rochei
C:Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 22-Oct-1999
C:Accession: JC2571; S34392
R:Perito, B.; Hanhart, E.; Irdant, T.; Iqbal, M.; McCarthy, A.J.; Mestromel, G.
A:Title: Characterization and sequence analysis of a Streptomyces rochei A2 endoglucanase
A:Reference number: JC2571; MUID:95011642
A:Accession: JC2571
A:Molecule type: DNA
A:Residues: 1-382 <PER>
A:Cross-references: EMBL:X79953; NID:g393391; PIDN:CAA52139.1; PID:g393392
A>Note: this cellulolytic strain was isolated from the gut of termites
C:Genetics:
A:Gene: egls
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation

```


OY	80	AAANTNTTAAK--MRVSYSTAAYMMDDIAAINVNGSGPGLTYLDAALSQOOGTPEV	137
Db	47	AASVRAAYGEOAFRLAQ-----DKIK--DANG-----YESA-----GQIPHA	87
OY	138	IEIYIDL-GRDCALASN-----GELPATAGL-QTYET-----QYDPIA	178
Db	88	VELTMAEPDYGEHTTSEASVYLMLEALYQGVOTDMAPLNHAFTMEKYMPIQSVDP	147
OY	179	SILSNPKYSSLRIYTIIEPDLPLNATYN-----MSIQACATAPVRYEGIEYALTKL	230
Db	148	NSFYNPSPPTAYAPDEFNHPSSTYQSLNSGISTGTDPIGAELKAT---YGNADVYQHMWL	203
OY	231	HAIPNVITYMDAAHSQ-WLQMPNNAGSYOEQVULNASI-----GVNGI	274
Db	204	ADVONITGFCATGACGCTLGPATGTSFINTQROGPQESWEHTYVQPCBEFGTKRNGY	263
OY	275	DGFVNTNANTYPLKEPEMTATQOVGOPIVES--ANFYQWMDIDEA-----	318
Db	264	LDLFTKDA5YA--KOMKTSASDADARAVEAYMAN--OMATEQKADVAATYAKAKM	319
OY	319	-DY-----AVDLYSLV-----AAGFBSISGLIDLTLRNGGNE-----	353
Db	320	GDIYRLYTLFKRYFKKICGTSPTCAAOQGBRAHYLLSWYMAWGATDTPSSGMAWRIGSSH	379
OY	354	-----PTGSTATVDYNTFVQNSITDLRQARGLQMNQNGAGL---	389
Db	360	AHFQYONPLAAMALSTDPKLTTPKSPTRAKDMA5M--QROLEY---TWLQASNGLIG	434
OY	390	-----GQPPASPTDPDPAHNLDAAYWIKRPGES--DGTSA-----ASDPTT	428
Db	435	ATNSMDGAYVQPRAGTFTFYGMGYTEAPYVYDPPSRMRGMQAMQGVQYAEIYASNAQ	494
OY	429	GKSDPMCDPYTTTSGYVLTNALPNSPIAGQFPAQFDOLYANARPAVPTSSSEPPP	488
Db	495	AKK-----IILDKWVP--WVYAN--ISTDQASMKYVSE	522
OY	489	PPSPASPPS	548
Db	523	LKMTGKPDPTMNAALPVTNP-----	541
OY	549	PSSSP	608
Db	542	-----GLTYE-----	546
OY	609	QLVNTGSSSYDLSTVYR---WYFRDGGST---LVYNCDMAAMCGNIRASFGSVN	660
Db	547	-VTSYGODVGAADTARALLFYAAKSGDASRDAKAKALLDAIWA-----NNQDPLGY-	597
OY	661	PATPATDYIQLST-----GGLAAGGSIGELQN--RVKKSQMSNFTENDSYSTNTTF	714
Db	598	SAVEREDGDKFDDTYVANGDGIYIPSGWTGIMPNDGVIKPGVSELDIRSFKKPDN--	654
OY	715	QDMTKVTVYVNGVLWMCETP-----SGTSPSTPPSPSPSP	749
Db	655	--WKSVOYTFLDG---GAEQOFRIHRFMAQTAVAGLADYALFDDGTT---TP-----	699
OY	750	SPSPGQDTPSPVPTGLVYTVGSSGSSVSLAMNASTDNVCAHYNRYRNGVLYGQPTVTSF	809
Db	700	-----DPTATVPTPTGLQACVYVSTETATISMTASTDTEHYTGDIYRGAKTKGATTTTSF	753
OY	810	TDGTGLAGTATYVVAADAGNTSABSTFVYATITTSSPSPSPPTPGTYVTDCTPEPNONG	869
Db	754	TDGTGLASTAAYVYRAFPDAAGNVSAPSAALVYTKRA-----TPSDTT-----APSYPA	802
OY	870	VTSVQGDXYRQVQTEHNSNAQCLITINTAGATVSTANPSGCTGAPATYIYKGCWH	929
Db	803	ITS-----SSS---TANSVYIGASATDN--AGSG--LAGY--DVRGA--	838
OY	930	GNCTKNYGMPIQISQIGASVATSMSTVOVSSGAY-----DVAADI-----WT	971
Db	839	-----TRAAQI--TALTFTDGLTASTAYEYTVARADYAGNVSAPSTAVYTKS	886
OY	972	NSPTPTTGO--PNGTEIMILMNSRGVOPFGSQTATGVTAAGHTMNWQGOQTSMKIIST	1029

Db	887	DTEPDTAPSPVAGLAA	-----			-----			-----	TVETSSVA-LTGN	-----	916
Qy	1030	VLPFGATGISNLDKAIFADAAARGSLNTSDYLDLVEAGFEIWMCGQGLGNSPSVS	YTS	1089								
Db	917	-----	ASIDTGGSLG	-----			-----	GIDYIKRATVGSITTA	STDT	949		
Qy	1090	GTSSEPTSPSEPTFPPESETPPSPSPSPSSBSSGVACRATVYVNSDMGSGF	ATV	1149								
Db	950	GLTATATVQYTVRATDNAGVNSAASALSVTTKTPDTG	-SCSVYNNAS	-WNSG	GFASV	1007						
Qy	1150	TVTNTGRATNGTVAMSEFGNQTVTNNWNTALTOGASVATNTLSYNNVLOPQSG	STFG	1209								
Db	1008	RITWNGFTTINGMSLGEFDLTAQGVQGGMSATWTOGSTVATINAPMWGLAPGQ	YDVG	1067								
Qy	1210	FNCSYSGTNA	PT-----	LSCT	1226							
Db	1068	FNCSHTGQNPFPASFTLNGASCT	1090									

RESULT 12
A39199
endoglucanase B (EC 3.2.1.-) - Cellulomonas fimi
C:Species: Cellulomonas fimi
C:Date: 08-Nov-1991 #sequence_revision 08-Nov-1991 #text_change 22-Oct-1999
C:Accession: A39199
R:Meinke, A.; Braun, C.; Glikes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.J
J. Bacteriol. 173, 308-314, 1991
A>Title: Unusual sequence organization in GenB, an inverting endoglucanase from Cellu
A:Reference number: A39199; MUID:91100298
A:Accession: A39199
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1045 <MEI>
A:Cross-references: GB:M64644; NID:g144415; PIDN:AAA23086.1; PID:g144416
C:Superfamily: fibronectin type iii repeat homology; bacterial cellulose-binding doma
C:Keywords: glycosidase; hydrolase
F:945-1045/Domain: bacterial cellulose-binding domain homology <BCB>
F:946-1044/Disulfide bonds: #status predicted

```

Query Match Summary: 8.9%; Score 580; DB 2; Length 1045;
Best Local Similarity 24.7%; Pred. No. 3.3e-17;
Matches 292; Conservative 101; Mismatches 356; Indels 432; Gaps 52;

QY 146 PGDR-----CAALASN-----GELPATAGL-----QYETQYIDPIASILSNP 184
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 181 PSSDVAETAAALAASSAIVLKGDSDPAVATLASHAKOLYTFADTYRGAVSDCVTAASAY 240

QY 185 K-YSSLRITYIIIEPSPLRNAVNMISIQCATNVPPEEGIEVALKRLAIPNVITYMDAA 243
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 241 KSMGSGYDELVMGCAVWLKATGDATYLAALAE-EVDKLTGENQST-----TRS 287

QY 244 HSGWLGMPRNNAAG-----YYQEVQKVLNA-SIGVNGIDGVTNANTPTLKE 289
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 288 YKMTIAMDKKQGTAYALLAMETGKKQKYVDANRWILDYMTVGNG-----QKVPSP--- 338

QY 290 PRMTATQOYGQPV-----ESAN-----FYQWNPD-----IDEADYAV- 322
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 339 -----GGQAVYLDMSGALRYAANTSEFVALVYSPDMWMTADRKARYHDFGRQJNTALVG 389

QY 323 -DLYSELVAGPSSSIGMLIDFLRMWGGSPEPTGSPSTATDVTNVSQKIDLRQHRGLM 381
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 390 DNPRESSIYVGF-----GANPTIAPHNHTANGSWLDSITTPPAQSRHLYL 433

QY 382 CNONGAGLGQPPQASFTDFPNNAHLDA---YVWIKPPGESDGTSAASDPTTGKKSDPMCDP 438
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 434 -----GALVGGP--GSPND---AYTDSRODYV-----ANEVATDYNAG----- 466

QY 439 TTTTSTGCVLTNMLNPSPILGQWFPQ--PDOLVANAARAVPISTSSPPPPPPSPSASPS 496
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 467 -TTSALARVVEEYGGPIPLSPPTPEPDGDCQLVEMLNQP----- 506

```

QY 497 PPSPPSPSSSPSPSPSSSPSPSPSPSPSSSPSPSSSPSPSPSPSPSPSPSPSSSPSPS 556
 Db 507 ----- 506
 QY 557 PSSPSPSPSPSPSPSSSPSPSPSPSPSPSPSPSPSSSPSPSPSPSPSPSPSPSPSPSPSS 616
 Db 507 -----PSGTTEV-----KAMIRNOSAPARS----- 528
 QY 617 SVDLSTVTVRWYFTRDG--GSSFLVYNCDMAAGCNGINIRASFGSVNPAIPADTYLQLSFT 675
 Db 529 ---LKNNAVRYWFTTIDGFASDVLTLSANKSE--CG---AQSGKGVSACTGIC-YVELSCV 579
 QY 676 GCTLAAGST--GEIQNRV-NKSDWSNFTETNDYSYG--TNTTFQDMTKVTVVYNGVLV 729
 Db 580 GQDIPGCGQSQRRIORLTCGPAGMN--PANDSYTGLTQTLAKASATILYXGSLV 636
 QY 730 MCTEESGTSPTSP 788
 Db 637 WKEPTPTGTTT-----DTPPTPTPATGTTVAGASLWMAASTDASC 680
 QY 789 VAHVWVY---NGVLGQPTVTSFTDGLAAGTAYTVTVAAVDAAGNTSAPSPVPAATT 844
 Db 681 VAGVLYLVNQGTQTLVETTAAAILRDLPGRATVSYVAKADVAGVSAASAATVFTT 740
 QY 845 TSPSPSPPTGTCTVTDCTPG--PNONGVTSVQGDXYRQVOTNEMSSAQOCLTINTAGMT 903
 Db 741 DT-----TGTEPTPTPGTPTVASAVTS-----TGATLAMA 770
 QY 904 VSTANFSGGTGAPATPYSIKKCGHWGNCCTKNVCMPIQISQISGAVTSMTTQVSSAY 963
 Db 771 PST-----GDPAV--SGIDVLNVGCTTTTVAQTTPVTYLSGLTPSTAYTVAAVRAK 820
 QY 964 DVAYDITMTNSP-----TTTGQPNGTETIMIMLSRCGVQFSGOTATGVTVAGHTNVMQGO 1020
 Db 821 NVAGDVSALSNAPVFTTAAPVDT-----VAP-----TYPGTVA----- 855
 QY 1021 QTSWKIIISYVLTLPGATSISNLDLKAIFADAAARGLNTSDYLLDVEAGFEIM-----OGG 1075
 Db 856 -----SNVATTCGATL-----TWTASTDSCG 875
 QY 1076 OGL-GSNFSVSYSVTSCTSSPTSPSPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 1120
 Db 876 SGLAVYEVLRVSGTQTIVASPTTATVALAGLTPTATVSYVRAKDGAGNVSAVSSPYTF 935
 QY 1121 TSPSSSGVACATVTVVNSDMGSGFTATVTVNTGSRATNGTVMASRFGNQTFVNTYNT 1180
 Db 936 TTLPTPTSPSCVTVVSTMS--MNVGFTGSVKITINTGTTPLT--WTLGFAPPSGQVYQGWSA 993
 QY 1181 ALTQSGASVTATNLSTYNNVYIQPGOSTTTFGFGSYSGTNAAP 1221
 Db 994 TMSOTGTTVTATATGLSMNATLQPGQSTGIDGFGNSHGTWTPN 1034

RESULT 13
 T34434
 hypothetical protein K06A9.1a - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T34434
 R:Geisler, C.; Gatlung, S.
 submitted to the EMBL Data Library, December 1996
 A:Description: The sequence of C. elegans cosmid K06A9.
 A:Reference number: 221525
 A:Accession: T34434
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2232 <GEI>
 A:Cross-references: EMBL:U08046; PIDN:AACT0890.1; GSPDB:GN00028; CESP:K06A9.1a
 A:Experimental source: strain Bristol N2; clone K06A9
 C:Genetics:
 A:Gene: CESP:K06A9.1a
 A:Map position: X
 A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/

Query Match	8.4%	Score 550	DB 2	Length 2232
Best Local Similarity	23.6%	Pred. No. 1.4e-15		
Matches	316	Conservative 190	Mismatches 509	Indels 326
				Gaps 51
16	GTRRPAIKRLRAGLAVAGVIAAIVLAQHPAIAATHVDNYPAGATFVNPMYMAQE	75		
Db	237	GTSSEPLRP-----SITSALPIA-----SSASSPS8AASSTTPVVLSSST	278	
Qy	76	VOSEANQTNATLAAKMRVSTYTAVMMDRIAINGVGBGLT-TYLDAA,SOOQGT	134	
Db	279	IGSSGTFPSSVAASPSTVGST-SGAASSSSATATSTIGSTITTPVGGSSITGSSST	337	
Qy	135	PEVIEIYITDLGRCOAAALASNGELPATPAAGIQETQYIDIPASILSPKISSLRITYI	194	
Db	338	PS-----ASSSSSGTM-STISG-STGSTVYVGGSS-----STASSTP	374	
Qy	195	IEPDLSLPNAVITNMSIOACTAAVPEYEOGIEVALTKLHAIPNYIYMDAHSGLMPNNA	254	
Db	375	IASSSPGSTVYVAPGSSST-----YSSSTPSA-----SSSSSGTMTNNGS	416	
Qy	255	SGYDEVQKVLASIGVNCIDGFVNTANTYTLKEPFMTATQVCGQPY----ESANFY	309	
Db	417	TGSTVYVAPVSSSTTG-----SSTPI-----ASSSSSGSTVYVGGSSSTYG	458	
Qy	310	QANPDIDEDADYAVDLYSLVLAAGFPSSILMLIDTLRNCGMGNEPGETATADVMTFVNO	369	
Db	459	SSTPSASS-----SAGTASTI-----SGSGTATITVPOSSSVGS----	495	
Qy	370	SKIDLRQHRGLMCNONGAGLGOPPOASPTDFPNALDAYWIKPGESEDGTSMAADPTTG	429	
Db	496	-----STOSAPSSPGTMTSTVSGCTGSTVYV-----PGSSTPSAPSSSP----	535	
Qy	430	KKSDMCPD-----TYTTSYGVLNMLPNBPICQWMPAPFDOLVANARAVPTSSS	482	
Db	536	---NPSSPASTGSTTTTIGSSSIIIVSYGSGTVSGSTGTSTOSTLASSATPGSSSTPS	592	
Qy	483	SPPPPPSP	538	
Db	593	SSSPQSPSSPAPNMGSTTPPSQTSQSPSPMANNPSSSTPTGSSQSTITPEGSTASPICS	652	
Qy	539	PS-----PSP	586	
Db	653	TGSTSVATEVYSQSTVSPGSSGLQSTNTPSSPSPSLSPSTSGMSTLITSEPSSTOSGG	712	
Qy	587	GLKVOYKKNMDSAPGNOIKPGLQVNTNGSSSDLTGVYRVYFTPDGSGSTLYVNCDMA	646	
Db	713	---AOSTLTTPPAPNSQSTSSLESSTSGAT-----TSSGSGITMTSPSQSS	756	
Qy	647	MCGGIRASFGSVNPATPTADTYLQ,LSFTGTLAAGS-----TGEIONRVKSDMS	698	
Db	757	---SVGSSQGSTSPAASST--TSGEMTSGSGTQTPGSSVSTSAALITSTQGSVSTNSPOS	810	
Qy	699	NTELENDISYGNITFQDMTKITYVNVGLVNGTPEPSGSPSPPTSPSPSPS-----	750	
Db	811	TYTRPSTVSGSTSS-----GSTVVGSTEAISTGSSVASSPAPSTSQNPPTSGSSSMI	866	
Qy	751	-----PSPGADV-----PSPVPT-----GLVYTVGSGSSVSLA	779	
Db	867	TQSPYPSQSTPVESTTPSPSPGPTLTLTSTSPSPSQSTTISTQSGTSPGISTISEMT	926	
Qy	780	WNASIDNNGVAHYNYRNGVLYGQ-TYTSFPTDGLAAGTATYVVAVD-----	828	
Db	927	SGGSGTQTPP-----STGSTVQGSTVSDSTSSG-----STVTVGSTEKSSSPISPTS	973	
Qy	829	-----AAGTASAPITVATTTTSPSPSPPT-----PTGTYTDCPGPNONG-VTS	872	
Db	974	QNTNPTSTSGSSMSQTPQSSQSTPVESTTSGAISTSSGSPPTLTLTSTSPSPSPSTICS	1033	
Qy	873	VOGDEYR-VQTNEMNSSAQOCLTINTATCAWTVSTANFSGGTGAPATVPSIYKCHMGN	931	
Db	1034	SGGTSNPNVYVSTIGSGSTSTPGSTGTVTKPSTVSSASGSGATATGSTEASTSGS--GS	1090	

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: August 29, 2002, 16:22:40 ; Search time 38.44 Seconds
(without alignments)
1236.930 Million cell updates/sec

Title: US-09-917-384-1
Perfect score: 6525
Sequence: 1 MERTQOGRNCRVORGTTRM.....GFNGSYSGTNAAPTLSCTAS 1228

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1655.5	25.4	872	1 GUXA_CELFI	P50401 cellulomona
2	744.5	11.4	1742	1 GUNA_CALSA	P22334 caldocellum
3	671.5	10.3	1331	1 MAMB_CALSA	P22333 caldocellum
4	641	9.8	438	1 GUNB_CELFI	P49075 agaricus bl
5	581.5	8.9	1090	1 GUNB_CELFI	P50899 cellulomona
6	580	8.9	1045	1 GUNB_CELFI	P26225 cellulomona
7	561	8.6	462	1 GUNB_FUSOX	P46336 fusarium ox
8	548	8.4	471	1 GUN2_TIRRE	P07987 trichoderma
9	543	8.3	562	1 GUN1_ACICE	P54583 acidotherma
10	519.5	8.0	1039	1 GUNB_CALSA	P10474 c endogluc
11	505	7.7	747	1 GUNB_CELFI	P50400 cellulomona
12	493.5	7.6	5179	1 MOC2_HUMAN	002817 homo sapien
13	488	7.5	1853	1 CIPA_CLOTM	006851 clostridium
14	472.5	7.2	772	1 CIPB_CLOTM	001866 clostridium
15	443	6.8	1367	1 AMW_H_YEAST	P08640 saccharomyc
16	416.5	6.4	474	1 VTP3_TTVIV	P19275 thermoprote
17	411	6.3	279	1 Y091_NPVOP	O10341 oryza psau
18	406.5	6.2	555	1 GRI_CHLRE	O9f096 chlamydomon
19	404	6.2	441	1 GUN2_THERU	P26222 thermomons
20	386	5.9	145	1 YCBA_PAEFLA	P29718 paenibacill
21	378	5.8	499	1 GUN2_BACSU	P10475 bacillus su
22	377.5	5.8	700	1 GUNA_PAEFLA	P29719 paenibacill
23	376	5.8	456	1 GUNA_MICBI	P26414 microbispor
24	375.5	5.8	499	1 GUN3_BACSU	P23449 bacillus su
25	370.5	5.7	499	1 GUN1_BACSU	P07983 bacillus su
26	369.5	5.7	1664	1 SLPI_CLOTM	O06852 clostridium
27	355	5.4	1140	1 YW6_YEAST	O04893 saccharomyc
28	353.5	5.4	504	1 GUNB_ERWCA	O05995 erythrina car
29	348.5	5.3	880	1 GUN4_THERU	P26221 thermomons
30	347.5	5.3	1229	1 N121_HUMAN	O9y223 homo sapien
31	346.5	5.3	1848	1 CIPA_CLOCT	P38058 clostridium
32	344	5.2	505	1 GUNV_ERWCA	O47096 erythrina car
33	339.5	5.2	484	1 GUX_CELFI	P07986 cellulomona

34	335	5.1	699	1 CH11_BACCI	P20533 bacillus ci
35	335	5.1	1970	1 RPBI_MOUSE	P08775 mus musculu
36	333	5.1	1970	1 RPBI_HUMAN	P24928 homo sapien
37	324.5	5.0	1802	1 HKR1_YEAST	P41809 saccharomyc
38	322.5	4.9	2090	1 N214_HUMAN	P35658 homo sapien
39	313.5	4.8	449	1 APG_BRANA	P40603 brassica na
40	313	4.8	268	1 NS20_MEDTR	P93329 medicago tr
41	313	4.8	1306	1 MSB2_YEAST	P32334 saccharomyc
42	312	4.8	1199	1 N121_RAT	P52591 rattus norv
43	310	4.8	986	1 GUN2_CLOSR	P23659 clostridium
44	310	4.8	1150	1 APMD_PIG	P12021 sus scrofa
45	309.5	4.7	914	1 GUX2_CLOSR	P50900 clostridium

ALIGNMENTS

RESULT	ID	STANDARD	PRT	872 AA.
1	GUXA_CELFI			
AC	P50401:			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Exoglucanase A precursor (EC 3.2.1.91) (Exocellulohydrolase A)			
DE	(1,4-beta-cellulohydrolase A) (CBP95).			
GN	CBHA.			
OS	Cellulomonas flm1.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.			
OX	NCBI_Taxid=1708;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 484;			
RX	MEDLINE=94344030; PubMed=8065260;			
RA	Meinke A., Gilkes N.R., Kwan E., Kilburn D.G., Warren R.A.J.,			
RA	Miller R.C., Jr.;			
RT	"Cellulohydrolase A (CbaA) from the cellulolytic bacterium			
RT	Cellulomonas flm1 is a beta-1,4-exocellulohydrolase analogous to			
RT	Trichoderma reesei CBH II."			
RL	Mol. Microbiol. 12:413-422(1994).			
RN	[2]			
RP	SEQUENCE OF 41-58.			
RX	MEDLINE=93209933; PubMed=8458833;			
RA	Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C., Jr., Warren R.A.J.;			
RT	"Cellulose-binding polypeptides from Cellulomonas flm1: endoglucanase			
RT	D (Cend), a family A beta-1,4-glucanase."			
RL	J. Bacteriol. 175:1910-1918(1993).			
CC	- FUNCTION: THIS ENZYME HYDROLYSES 1,4-BETA-D-GLUCOSIDIC LINKAGES OF			
CC	CELLULOSE. WEAK ACTIVITY AGAINST CARBOXYMETHYLCELLULOSE, BACTERIAL			
CC	MICROCRYSTALLINE CELLULOSE AND BARLEY BETA-GLUCAN. HAS ALSO WEAK			
CC	ENDOGLUCANASE ACTIVITY. HYDROLYSES GLUCOSIDIC BONDS WITH INVERSION			
CC	OF ANOMERIC CONFIGURATION.			
CC	- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages			
CC	in cellulose and cellotriose, releasing cellobiose from the non-			
CC	reducing ends of the chains.			
CC	- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.			
CC	- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN			
CC	(CBD).			
CC	- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL			
CC	HYDROLASES).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: L25809; AAC36898.1; -			
DR	HSSP: P07986; IEXG.			
DR	InterPro: IPR001919; CBD_2.			

DR InterPro: IPR003961; FN.III.
 DR InterPro: IPR003962; FN.III repeat.
 DR InterPro: IPR001524; Glyco_hydro_6.
 DR Pfam: PF00553; CBD_2; 1.
 DR Pfam: PF00041; fn3; 3.
 DR Pfam: PF01341; Glyco_hydro_6; 1.
 DR PRINTS: PR00014; ENTPERTII.
 DR PRINTS: PR00733; GLHYDRLASE6.
 DR ProDom: PD003733; Glyco_hydro_6; 2.
 DR SMART: SM00060; FN3; 3.
 DR PROSITE: PS00561; CBD_BACTERIAL; 1.
 DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
 DR PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
 KM Cellulose degradation: Hydrolase; Glycosidase; Repeat; signal.
 FT SIGNAL 1 40
 FT CHAIN 41 872
 FT DOMAIN 41 477 EXOGLUCANASE A.
 FT DOMAIN 478 563 CATALYTIC.
 FT DOMAIN 573 664 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 673 768 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 769 872 FIBRONECTIN TYPE-III 3.
 FT ACT_SITE 188 188 CELLULOSE-BINDING (BY SIMILARITY).
 FT ACT_SITE 188 188 PROTON DONOR (BY SIMILARITY).
 FT DISULFID 140 410 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 374 428 BY SIMILARITY.
 FT DISULFID 770 869 BY SIMILARITY.
 SQ SEQUENCE 872 AA; 89300 MW; 7883B407F995533B CRC64;

Query Match 25.4%; Score 1655.5; DB 1; Length 872;
 Best Local Similarity 33.1%; Pred. No. 5.7e-59;
 Matches 408; Conservative 125; Mismatches 300; Indels 399; Gaps 28;

QY 23 ISKRLRGVLAGAVSTAIATPLA--MHPALIAATHVDNPAAGTFYVNPYMAEVOSEA 80
 DB 10 VRRKVRRA-VATAAATATLAVVPLTLTATSASAPVHDNPAAGVQVNPYMAASVYNA 68
 QY 81 ANCT-NATTLAKMRVSTYSTAVVMDRIAINGVNGSGFTTYLDALISQOQGT-TPEVI 138
 DB 69 GROSADALAAKMTTAVGPTAVVMDRIASALTGNADNGKLFHLDNAVQAQKAAGVPLVF 128
 QY 139 EIVTYDLPGRDCAALASNGELPATVAGLQYETQYIDPIASILSNPKYSLRIYITIEP 198
 DB 129 NLVYDLPGRDCAALASNGELPATVAGLQYETQYIDPIASILSNPKYSLRIYITIEP 188
 QY 199 SLRPAVYVMSIQACATAVPYEGIEVALTKLHAIPVYITMDAASHGWLCPNNA---S 255
 DB 189 SLRPAVYVMSIQACATAVPYEGIEVALTKLHAIPVYITMDAASHGWLCPNNA---S 248
 QY 256 GYVQEVQKVLNASIGVNGIDGFTNTANTYPLKEPFMT-ATQOYGGQGVESANFYQWNP 314
 DB 249 TLRAEVAK--STTAGFASIDGFSVDVANTYPLEEPLSDSLTINNPIRSKFEYEMFD 306
 QY 315 IDEADYAVDLYSRLVAGFPSSIGMLIDTLNGVNGGNEPTGPSTANDVNTFVNSKIDL 374
 DB 307 FDEIDYTAHMRRLVAGFPSSIGMLIDTLNGVNGGNEPTGPSTANDVNTFVNSKIDL 366
 QY 375 ROHRLGLCNGAGLQGPPOASPTDFPNAHLDAYVTKPGEISGTSASAPPTGKSDP 434
 DB 367 RVHGRAMCNPAGIGRFPPEPGVAAASHDAFWIKPPEISGASATDIPNDGKRDR 426
 QY 435 MCDPTVYTS--YGLVTNALPNSPIAGOMEPAPQDQVLANAPAVPTSSSSPPPPPPPS 492
 DB 427 MCDPTVYSPKLNOLTAATVAPLAGOMFEQFVTLVNAIPVIGGT-----PVEDL 479
 QY 493 ASPSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSS 552
 DB 480 VAPF----- 483
 QY 553 PSPSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSS 612
 DB 484 -----VPTGLTAGT 492

QY 613 TGSSVDLSTVTVRWFTRDGGSTLVYNCMDAAMGCGNIRASGVSYPATPATYIQL 672
 DB 493 TTATSVPLS-----WTA-----STDNVATGIDVYRG 520
 QY 673 SFTGGTLAAGSGTGEIONRVKSDMSNFTETNDYSYGTNTTFQDWTVYVYVNGVLWGT 732
 DB 521 TLVGTAA-----TSYVLTGLTPATAV--SETVRAK 549
 QY 733 EPSGTSPTSPSPSPSPSPSGDVTPSPVPTGLVYTVGSSSVSLAMNASTDNV---GV 789
 DB 550 DAAGNVSASASAAAATQSGTIVTDTPAPSPAGLACTTTTTPVPLMWTASTDAGSGV 609
 QY 790 AHVYVYRNGVAVGQPTVSTFDTLAAGTATYTVAAVDAAGNTSAPSTPYAT----- 843
 DB 610 AGVELKGTIVYGTITATSTVYTGTLTACTTYSFVSRAKDVAGNTSASAAVSATTOGT 669
 QY 844 --TTPSP--PSPPTPTGTVDCTPGPNONGVTSVQDEYRVQTNEMNSAQOCLTINATG 900
 DB 670 VDTAPSPVPTGLTAGTTTSSVP-----LT----- 694
 QY 901 AMYVSTANFSGGTGAPATYTSITKGMGNCCTKNVMPVQISQISAVTSMTTOVSS 960
 DB 695 -WTASTDN-AGSGVA----- 708
 QY 961 GAYDAVDIWNSTPTTGTGPNGTETIMILNLSRGVQPFQSGATAGVYVAGHTNVMQ 1020
 DB 709 -----GVEFNGTTRVAT-----VTSTSTVYTG----- 731
 QY 1021 QTSWKIISYVLPATGATISLNLKAIFADAAARSLNTSDYLDLVEAGFEIMOGGIGS 1080
 DB 732 -----LADTAYSEYVAKKDVAGVNSA----- 754
 QY 1081 NSFVSYSVSGTSPTSPSPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 1140
 DB 755 --SAAVSARTQAT-----SG-GCTVYTSAS- 778
 QY 1141 MGSFATVYVNTGSRATNGTVAAMSPGQNTYNTVNTALTOGASVATNTSYNVI 1200
 DB 779 WNTGFTGVEKKNNTALNWTGLGFSRADQKXSGSASAMSQSGTAVAKNAPWNTL 838
 QY 1201 QPGQSTTFGNGSYSGTNAAPT-----LSCT 1226
 DB 839 AAGSSVSIGFNGTHNGTAPTAFTLNGVACT 870

RESULT 2
 GUNA_CALSA STANDARD; PRT; 1742 AA.
 AC P22534;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)
 GN CELA.
 OS Caedocellum saccharolyticum (Caedocellulolusruptor saccharolyticus).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Thermocaneorobacter group; Caedocellulolusruptor.
 OX NCBI_TaxID=44001;
 RN [1]
 RP MEDLINE-95336703; PubMed-7612247;
 RA Te'O V.S., Saul D.J., Bergquist P.L.;
 RT "celA, another gene coding for a multidomain cellulase from the
 extreme thermophile Caedocellum saccharolyticum."
 RL Appl. Microbiol. Biotechnol. 43:291-296(1995).
 RN [2]
 RP SEQUENCE OF 1516-1742 FROM N.A.
 RX MEDLINE-91247819; PubMed-2039230;
 RA Leuthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
 RT "Cloning, sequence analysis, and expression in Escherichia coli of a
 gene coding for a beta-mannanase from the extremely thermophilic
 bacterium 'Caedocellum saccharolyticum'.
 RT

```

RL  Appl. Environ. Microbiol. 57:694-700(1991).
CC  -1- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
CC  ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL
CC  DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
CC  CELLULOSE.
CC  -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC  linkages in cellulose.
CC  -1- PRM: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL
CC  SITE FOR PROTEOLYSIS.
CC  -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC  E (FAMILY 9 OF GLYCOSYL HYDROLASES).
CC  -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
CC  L (FAMILY 48 OF GLYCOSYL HYDROLASES).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  DR  EMBL; L32742; AAA91086.1; -.
CC  DR  EMBL; M36063; AAA72860.1; -.
CC  DR  EMBL; L01257; -. NOT_ANNOTATED_CDS.
CC  DR  PIR: A43745; A43745.
CC  DR  HSSP; P26221; 1TF4.
CC  DR  InterPro; IPR001956; CBD_3.
CC  DR  InterPro; IPR000556; Glyco_hydro_48.
CC  DR  InterPro; IPR001701; Glyco_hydro_9.
CC  DR  Pfam; PF00942; CBD_3; 3.
CC  DR  Pfam; PF02011; Glyco_hydro_48; 1.
CC  DR  Pfam; PF00758; Glyco_hydro_9; 1.
CC  DR  PRINTS; PR00844; GLHYDRASE48.
CC  DR  ProDom; PD001947; CBD_3; 2.
CC  DR  ProDom; PD011903; Glyco_hydro_48; 1.
CC  DR  PROSITE; PS00592; GLYCOSYL_HYDROL_F9.1; 1.
CC  DR  PROSITE; PS00698; GLYCOSYL_HYDROL_F9.2; 1.
CC  K4  Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
CC  FT  SIGNAL 1
CC  FT  CHAIN 23
CC  FT  DOMAIN 24 1742
CC  FT  DOMAIN 24 642
CC  FT  DOMAIN 643 700
CC  FT  DOMAIN 701 857
CC  FT  DOMAIN 858 903
CC  FT  DOMAIN 904 1060
CC  FT  DOMAIN 1061 1112
CC  FT  DOMAIN 1113 1742
CC  FT  ACT_SITE 396 396
CC  FT  ACT_SITE 434 434
CC  FT  ACT_SITE 443 443
CC  FT  CONFLICT 1545 1545
CC  SQ  SEQUENCE 1742 AA; 193696 MW; 3F0699A213EED07 CRC64;

```

[illegible]

Dd	315	GYNBERIKYTPKGLAMLDQMSLRVATTTAFLAFVSDMSGCPTEKKETRYRKFSGSODIY	374
Oy	226	AL-----TKLHAPNVYIYMDAAHSGN-----LGWPNNASGYO	259
Dd	375	ALGSTRGSFVVGFCGTNPBRPHHRHTAHSSMADQSISPYRHRTLYGALVGGGSDSYTD	434
Oy	260	EVOQKVLASIGVANGIDGFVTNTANY-----TPLKEPR-----	292
Dd	435	DISNYVNVEACDYNAGFEVALAKMYLLYGCPIDPFKAIEPTPDEDFEENGJINASTGN	494
Oy	293	-----TATOOGGVPSANFYOMNDPIDBADVAPLYSLVLAAGFPSSICMLIDTLANG	347
Dd	495	FIEIKAIYNNOSGHPARATIKLKR-----YFDL-SELIKAG-----	532
Oy	348	MGPNEBPGSTADTVTFVNOQKIDLROHRLCMONAGAGOPPOASPDPFNMAHLDA	407
Dd	533	--SPNOULT-----ISTNYNQ-----CAKVSQP-----	552
Oy	408	YVW-----IKPPGESDGTSAASDPITGCKSPMKCDPPTYTSYGLVTAL	451
Dd	553	YVMOSSRNIIYYILVDFTGLIYPGGDKKKEYOFRIAPQWOWONDNSDYFODIKCVS	612
Oy	452	PNSDIAGOMPPAQPDOL-VANARPVPYTRSSSPPPPSPASPSpspspspspspspsps	510
Dd	613	SGSVYTKKIYPLDIEDIKWGEERG-----TSGVPT	644
Oy	511	PSPSspSP	570
Dd	645	PLAVPTPLPPP-----TLPATPTPTPTPVTPPTPVATPTPTPTPTPTPTPT	688
Oy	571	PSSPSPSPSPSP--VSGGLKVQYKNNDAPGDNOIKPLQLVNTGSSVDSLTVTRYWF	629
Dd	689	PTVPPTPTPATSPATSOIKVLYLANKEFTNSTWTIRPWKLVVNSSSSSIDLSRVTIERYX	748
Oy	630	TRODGSSSTIVYNCOMMAMGCCNIRASFGSNPATPIADTYILOLSTTCGR--LAAAGSGE	687
Dd	749	TVDERROSALS-DMAOIGASNTEFKVKLISSSVSCADYLLEIGRSKGAGLOPKKDGE	807
Oy	688	IQNENKSDMSNFTETDSY--CTNTTFODMTVYVYVNGVLMGTERTSGTSPSPSPSPS	746
Dd	808	IQIFNDDMSNTYNOGDNMHIOSMTSYGENEVTAVIDGVLMOEBSGTTIPATISPT	867
Oy	747	PSPSPSFGDTPSPVPTGLVTVTGSSSVSLAMNASTDNGVAHYNYRNGVLGOPTV	806
Dd	868	VY-----VTPPTPTPT-----PTV	880
Oy	807	TSFTDTGLAGTATTYVAADVADAAGTNSAPSTPVATTTSPPSPPTCTTYTOCTBPBN	866
Dd	881	T-----PPTPVNT-----PPTPVPTSPSPS-----	902
Oy	867	ONGVTVSOGDEVRY-----OTNEMNSSAOOCL-TINTATGAMTVSTANPSGGTGAPATY	920
Dd	903	-----TPATGOIKVLYANKKETNSTWTIRPWKLVVNSSSSSIDLSRVYTR-----	948
Oy	921	PSIYKGGMGNCCTTKNVGMPIQIOIGSAVTSNSTOVSSGAYDVADYIMTNSTEPITGO	980
Dd	949	-----YW-----TVDERROSALSIDMA-----QIGASNTEFK-FVWLSSSVGA	987
Oy	981	PNGREIMIMLNBSRGVOPRESSOTATGVYAAGHW-NVMGCOOTSUKIISYLTLPATIS	1039
Dd	988	DYYLEIG-FKSGAQOLOPGHDTEGIOIRFNKDOMSYNONGNDMSW-----IQSMITYG	1039
Oy	1040	NLDIKATFADAAANGSLNTSDYLLDYEAGEFEIMOGGOLGSMFSVYSVTSCHTSPTSPS	1099
Dd	1040	ENEKYTAIID-----GYLVN-GQEPSGATPAPATPVPTPVTPPTPAPA	1080
Oy	1100	PTPTSPPTPTPSP--SETPSPSPSTSPSPS	1127
Dd	1081	PTPTATPTPTPTPTPTPVTPPTPVAPTPTPTPSPST	1110

RESULT 3
MANB_CALSA

RESULT	3
MANB_CALSA	

ID	NAME	STANDARD	PT	AA
AC	P22533			
AD	MAMB_CALSA			
DT	01-AUG-1991	(Rel. 19, Created)		
DT	01-JUL-1993	(Rel. 26, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Beta-mannanase/endorganase A precursor [includes: Mannan endo-1,4-beta-mannosidase A (EC 3.2.1.78) (Beta-mannanase) (Endo-1,4-mannanase); Endo-1,4-beta-glucanase (EC 3.2.1.4) (Cellulase)].			
GN	MAMB.			
OS	Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Thermoaerobacter group; Caldicellulosiruptor.			
OX	NCBI_taxonomy:44001;			
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=9311939; PubMed=1476429;			
RX	Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;			
RA	"The beta-mannanase from 'Caldocellum saccharolyticum' is part of a			
RT	multidomain enzyme.";			
RL	Appl. Environ. Microbiol. 58:3864-3867(1992).			
RN	[2]			
RP	SEQUENCE OF 1-346 FROM N.A.			
RX	MEDLINE=91247819; PubMed=2039230;			
RA	Luthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;			
RT	"Cloning, sequence analysis, and expression in Escherichia coli of a			
RT	gene coding for a beta-mannanase from the extremely thermophilic			
RT	bacterium 'Caldocellum saccharolyticum'.";			
RL	Appl. Environ. Microbiol. 57:694-700(1991).			
CC	-1- FUNCTION: DEGRADATION OF HEMICELLULOSES. THE SECOND MOST ABUNDANT			
CC	POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH			
CC	MANNAASE AND ENDOLUCANASE ACTIVITIES.			
CC	-1- CATALYTIC ACTIVITY: Random hydrolasts of 1,4-beta-D-mannosidic			
CC	linkages in mannan, galactomannans, glucomannans, and			
CC	galactoglucomannans.			
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic			
CC	linkages in cellulose.			
CC	-1- MISCELLANEOUS: THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREES			
CC	CELSIUS.			
CC	-1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY			
CC	A (FAMILY 5 OF GLYCOSYL HYDROLASES).			
CC	-1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY			
CC	J (FAMILY 44 OF GLYCOSYL HYDROLASES).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL; L01257; AAA71987.1; -			
DR	EMBL; M36063; AAA72861.1; -			
DR	PIR; B43745; B43745.			
DR	PIR; A48954; A48954.			
DR	HSSP; Q06851; INBC.			
DR	InterPro: IPR001547; Glyco_hydro_F5.			
DR	InterPro: IPR001547; Glyco_hydro_F5.			
DR	Pfam; PF00942; CBD_3; 2.			
DR	Pfam; PF00150; cellulase; 1.			
DR	ProDom; PD001947; CBD_3; 2.			
DR	ProSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.			
DR	HydroLase; Glycosidase; Cellulose degradation; Signal;			
KW	Multifunctional enzyme.			
FT	SIGNAL	1	411	POTENTIAL.
FT	CHAIN	42	1331	BETA-MANNAASE/ENDOLUCANASE A.
FT	DOMAIN	326	325	CATALYTIC (MANNAASE ACTIVITY).
FT	DOMAIN	462	361	PRO/SER/THR-RICH (PT BOX).
FT	DOMAIN	518	518	SUBSTRATE-BINDING (POTENTIAL).
FT	DOMAIN	519	564	PRO/SER/THR-RICH (PT BOX).
FT	DOMAIN	565	720	SUBSTRATE-BINDING (POTENTIAL).
FT	DOMAIN	721	780	PRO/SER/THR-RICH (PT BOX).
FT	DOMAIN	781	1331	CATALYTIC (ENDOLUCANASE ACTIVITY).

FT	ACG SITE	162	162	PROTON DONOR (BY SIMILARITY).
FT	ACG_SITE	257	257	NOCEPHOLITE (BT SIMILARITY).
FT	CONFLICT	338	338	T -> P (IN REF. 2).
FT	CONFLICT	340	346	TPPTPT -> RQHQHQ (IN REF. 2).
FT	SEQUENCE	1331 AA;	146992 MW;	FFBCA51BB8BF0E0 CRC64;
Query Match				
Best Local Similarity		10.3%;	Score 671.5;	DB 1; Length 1331;
Matches 243;		Conservative 145;	Mismatches 305;	Indels 341; Gaps
QY	240	MDAAHSGMLGPNN-----ASGY-----VOEYQKV--LMSISGVNGI-----	274	
DB	67	LDPLALRGIRSMGMSYRVVLNSGYRWTKIPASEVANIISLSLGFKAIIIEVHDTGYG	126	
QY	275	-DGVVTIAN-----YPLK-----EPPATIQVGQGVESANFYQNNPIDEDYAVD	323	
DB	127	EDGAACGSLAAVEYWEIKESVLDGNDFYII--NIGNEPYGNNTQNNVDTKNAIKA--	182	
QY	324	LYSLRVLAPESSIGMLIDLRNG--WGPNEPTGS-----TA	360	
DB	183	-----LRAGRKHL--MVDAIPMNGQMSMTMDNMQSIHEADPLNLVFSIMHYVITA	236	
QY	361	TDVTEFYNSQKIDLRHGRIGMCNAGAGLGPQASPT--DFPAHLDAYYVIRKPGES	417	
DB	418	DGTSAADPTTGKKSDPMC-DPTVTTSYGLNALPNSPIA--GQFPFAQFDLVANMARA	475	
QY	271	AIYVAKAQYIKGLFSWMCNCSSTVGYLDMVNNWMDNNPTPGQWYKTN-----A	320	
DB	476	VPTSTSSPPPPPPSPASPSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSS	535	
QY	321	IGT-----SSPTPTS	331	
QY	536	SSSPSPSPSPSPSSPSPPSSPSPPSPSPSPSSPSPPSSPSPPSSPSPPSSPSPPSS	594	
DB	332	TVLTPTPTPTPTPTPTVATP-----PTPTPTPTPTPTPTPTPTPTPTPTPTPTPT	373	
QY	595	NDSPAGNOIKRPGIOLATVNTGSSSVDSITVYRWTFRDOGSSSTLYVNCMDAMAGCINRA	654	
DB	374	KETNSTNTIRPMLKAYVNSGSSSIDLSRPTIRKWTYVDSERQASNIS--DMAQIGASNTF	432	
QY	655	SFGSVNATPTADTYLLOLSTGGT--LAAGSGTGLQNRVYKSDMSNFTENDISY--GTN	711	
DB	433	KFYVLISSVSGADYYLEIGKSGAGOLQKQDGTGELOMFNFDMSNNOGNDMSWIOSM	492	
QY	712	TTFDQMTKVVYVYVNGVYVWGTSEGTSPSPSPSPSPSPSPSGGVDTPPSVPTGLVGVY	771	
DB	493	TSYENKRYATYIDGYLVWQEPSCGATPAPATPTPTPT-----VTP-----	536	
QY	772	SGSSVSILAMNASTDNGVAHNYRVNGVLVGPQPTVTSFTDTGLAAGTAYTVVAADAG	831	
DB	537	-----PTPTV-----	541	
QY	832	NTSAPSTPVATTTSPSPSTPTGTIVTDCIPBNONGTYSVQDEHY-----QTNEMN	886	
DB	542	-----PTPTVAT--PTPTPTPTPTPS-----TPATGQIVLVANKTSTNPTT	583	
QY	887	SSAQOCL--TNTATGAMTVSTANFSGGTGAPATYPSIKGCHMGNCTTKNMGPIQISQ	945	
DB	584	NTIRPMLKAYVNSGSSSIDLSRVTIR-----YW-----YVDSER	617	
QY	946	IGSAVTSWSTQVSSGAYDAVAYIMTNSPTPTTGGPNGTEIMIMILNSRQGVQFSSQTAT	1005	
DB	618	AQASISDMA-----QIGASNTFK--FYKLSSVSGADYYLEIG--FKSGAGOLQPKGDTEI	671	
QY	1006	GVTYAGHTW--NVWOGQGTSKRIISVYLTPGATYSISMLDKAIFADAAAAGSLMSTDYLLD	1064	
DB	672	QIRPNKSDMSNYNOGNDMSH-----IOSMTSYGENEVTVYIID-----	709	
QY	1065	VEAGFEIMVGGQIGSNSFSFVSVTSGTSSPTSPSPPTPTPTPTPTPTPTPTPTPTPTSP	11244	

[illegible]

Query Match	8.6%;	Score 561;	DB 1;	Length 462;
Best Local Similarity	32.5%;	Pred. No. 4.8e-16;		
Matches 150;	Conservative 71;	Mismatches 167;	Indels 74;	Gaps 16;
17 TTRPAISKRRLAGVLAGAVSTAASIVPLAMOHPAIATHTVDNPFYAGATFEVNPWAOEV	17 TTRPAISKRRLAGVLAGAVSTAASIVPLAMOHPAIATHTVDNPFYAGATFEVNPWAOEV	17 TTRPAISKRRLAGVLAGAVSTAASIVPLAMOHPAIATHTVDNPFYAGATFEVNPWAOEV	17 TTRPAISKRRLAGVLAGAVSTAASIVPLAMOHPAIATHTVDNPFYAGATFEVNPWAOEV	17 TTRPAISKRRLAGVLAGAVSTAASIVPLAMOHPAIATHTVDNPFYAGATFEVNPWAOEV
70 TAAAGSSTATRTTATGTGSSITAGGSVSTA--PPAAS--DNPPYAGVDLANNYYRSEV	70 TAAAGSSTATRTTATGTGSSITAGGSVSTA--PPAAS--DNPPYAGVDLANNYYRSEV	70 TAAAGSSTATRTTATGTGSSITAGGSVSTA--PPAAS--DNPPYAGVDLANNYYRSEV	70 TAAAGSSTATRTTATGTGSSITAGGSVSTA--PPAAS--DNPPYAGVDLANNYYRSEV	70 TAAAGSSTATRTTATGTGSSITAGGSVSTA--PPAAS--DNPPYAGVDLANNYYRSEV
77 OSEAA--NQNATLTAAKRRVYSTSTANWMDRIALINVCNGPGITTYLDALALSOQOCT	77 OSEAA--NQNATLTAAKRRVYSTSTANWMDRIALINVCNGPGITTYLDALALSOQOCT	77 OSEAA--NQNATLTAAKRRVYSTSTANWMDRIALINVCNGPGITTYLDALALSOQOCT	77 OSEAA--NQNATLTAAKRRVYSTSTANWMDRIALINVCNGPGITTYLDALALSOQOCT	77 OSEAA--NQNATLTAAKRRVYSTSTANWMDRIALINVCNGPGITTYLDALALSOQOCT
124 MMLAIPKUSGAKATATAAKRADVPSPQ--WMQTYDHISLME-----DTLADLRKANKAG-174	124 MMLAIPKUSGAKATATAAKRADVPSPQ--WMQTYDHISLME-----DTLADLRKANKAG-174	124 MMLAIPKUSGAKATATAAKRADVPSPQ--WMQTYDHISLME-----DTLADLRKANKAG-174	124 MMLAIPKUSGAKATATAAKRADVPSPQ--WMQTYDHISLME-----DTLADLRKANKAG-174	124 MMLAIPKUSGAKATATAAKRADVPSPQ--WMQTYDHISLME-----DTLADLRKANKAG-174
134 TPEVEIIVYDLPGRDCAALASNGELPAPTAALQOTETQYIDPIASILSNPYSSLRIVT-193	134 TPEVEIIVYDLPGRDCAALASNGELPAPTAALQOTETQYIDPIASILSNPYSSLRIVT-193	134 TPEVEIIVYDLPGRDCAALASNGELPAPTAALQOTETQYIDPIASILSNPYSSLRIVT-193	134 TPEVEIIVYDLPGRDCAALASNGELPAPTAALQOTETQYIDPIASILSNPYSSLRIVT-193	134 TPEVEIIVYDLPGRDCAALASNGELPAPTAALQOTETQYIDPIASILSNPYSSLRIVT-193
175 GKVAQGFVYDLPNNDCAAAASNGEISLDKDKANRYKA-TYAKIKIGILN--YSDTKYVL-231	175 GKVAQGFVYDLPNNDCAAAASNGEISLDKDKANRYKA-TYAKIKIGILN--YSDTKYVL-231	175 GKVAQGFVYDLPNNDCAAAASNGEISLDKDKANRYKA-TYAKIKIGILN--YSDTKYVL-231	175 GKVAQGFVYDLPNNDCAAAASNGEISLDKDKANRYKA-TYAKIKIGILN--YSDTKYVL-231	175 GKVAQGFVYDLPNNDCAAAASNGEISLDKDKANRYKA-TYAKIKIGILN--YSDTKYVL-231
194 IIEPDSLPAVTNMSIQACATAPVPEYEGIEVALTKLHAIPNVYIYMDAAHSGWLGWPN-253	194 IIEPDSLPAVTNMSIQACATAPVPEYEGIEVALTKLHAIPNVYIYMDAAHSGWLGWPN-253	194 IIEPDSLPAVTNMSIQACATAPVPEYEGIEVALTKLHAIPNVYIYMDAAHSGWLGWPN-253	194 IIEPDSLPAVTNMSIQACATAPVPEYEGIEVALTKLHAIPNVYIYMDAAHSGWLGWPN-253	194 IIEPDSLPAVTNMSIQACATAPVPEYEGIEVALTKLHAIPNVYIYMDAAHSGWLGWPN-253
232 VIEPDSLAVLTNLNVDCAKAEASVYKELTYVAIKELN-LPNUVMYTLDRGHGWLGMWPN-290	232 VIEPDSLAVLTNLNVDCAKAEASVYKELTYVAIKELN-LPNUVMYTLDRGHGWLGMWPN-290	232 VIEPDSLAVLTNLNVDCAKAEASVYKELTYVAIKELN-LPNUVMYTLDRGHGWLGMWPN-290	232 VIEPDSLAVLTNLNVDCAKAEASVYKELTYVAIKELN-LPNUVMYTLDRGHGWLGMWPN-290	232 VIEPDSLAVLTNLNVDCAKAEASVYKELTYVAIKELN-LPNUVMYTLDRGHGWLGMWPN-290
254 ASGVQVEQKVLNASTGVNGIDGFPYTNANTYPLKEPFTATQOYGGQGVESANFYQWNP-313	254 ASGVQVEQKVLNASTGVNGIDGFPYTNANTYPLKEPFTATQOYGGQGVESANFYQWNP-313	254 ASGVQVEQKVLNASTGVNGIDGFPYTNANTYPLKEPFTATQOYGGQGVESANFYQWNP-313	254 ASGVQVEQKVLNASTGVNGIDGFPYTNANTYPLKEPFTATQOYGGQGVESANFYQWNP-313	254 ASGVQVEQKVLNASTGVNGIDGFPYTNANTYPLKEPFTATQOYGGQGVESANFYQWNP-313
291 IGPAAKLVAQIYTKDAGKBPBRVGLVATNVSNN-----GWLKSTRKPTDENSEP-337	291 IGPAAKLVAQIYTKDAGKBPBRVGLVATNVSNN-----GWLKSTRKPTDENSEP-337	291 IGPAAKLVAQIYTKDAGKBPBRVGLVATNVSNN-----GWLKSTRKPTDENSEP-337	291 IGPAAKLVAQIYTKDAGKBPBRVGLVATNVSNN-----GWLKSTRKPTDENSEP-337	291 IGPAAKLVAQIYTKDAGKBPBRVGLVATNVSNN-----GWLKSTRKPTDENSEP-337

QY 314 DIEDADYADVLYSRVAVAGPPSSIGMLIDPLRMGSGPNEPTGSTATDVNTFVNOSKID 373
DB 338 NYDEORY-INAFAVLAEOGMSNKFVDOGRSG---KOPTG----- 375
QY 374 LRORGLMCNONGAGLGPQAPSPDPFNAHLAVYVIRKPGESDGTSAASDPPTGKSD 433
DB 376 -QAAGQCMCAKAGTGFELRSTNGD---ALADLFVWVKGGESDGTG---DTSARYD 427
QY 434 PMCDPTVTSYGVLTNALPNSPIAGWFPAPQDOLVANARPA 475
DB 428 YHCG-----LDDALRKAPEAGTWFOAYFEQLDNANPS 460
RESULT 8
GUX2_TRIRE STANDARD; PRT; 471 AA.
AC P07987;
ID GUX2_TRIRE
DT 01-ANG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Exoglucanase II precursor (EC 3.2.1.91) (Exocellulohydrolase II)
GN CBH2
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VTT-D-80133;
RA MEDLINE=87248061; PubMed=3596237;
RT Teeri T.T., Lehtovaara P., Kauppinen S., Salovuori I., Knowles J.;
RT "Homologous domains in Trichoderma reesei cellulolytic enzymes: gene
RL sequence and expression of cellobiohydrolase II."; Gene 51:43-52(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-QM9414 / ROT C-30;
RA Chen C.M., Gritzall M., Stafford D.W.;
RT "Nucleotide sequence and deduced primary structure of
RL cellobiohydrolase II from Trichoderma reesei."; Biotechnology 5:274-278(1987).
RN [3]
RP SEQUENCE OF 25-44.
RA Faegerstaem L.G., Pettersson L.G.;
RT "The 1,4-beta-glucan cellobiohydrolases of Trichoderma reesei QM
RL 9414."; FEBS Lett. 119:97-100(1980).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=90333255; PubMed=2377893;
RA Rouvinen J., Bergfors T., Teeri T.T., Knowles J.K.C., Jones T.A.;
RT "Three-dimensional structure of cellobiohydrolase II from Trichoderma
RL reesei."; Science 249:380-386(1990).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=97029636; PubMed=8875646;
RA Koiwala A., Reinikainen T., Ruohonen L., Valkeajärvi A.,
RA Claeysseens M., Teleman O., Kleywegt G.J., Szardenings M., Rouvinen J.,
RA Jones T.A., Teeri T.T.;
RT "The active site of Trichoderma reesei cellobiohydrolase II: the role
RL of tyrosine 169."; Protein Eng. 9:691-699(1996).
CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC (1) EXOCELLULOBIODHROLASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC (2) EXOCELLULOBIODHROLASES THAT CUT THE DISSACCHARIDE CELLULOSE
CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-

CC reducing ends of the chains.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: T. REESEI PRODUCES TWO DIFFERENT
CC EXOCELLULOBIODHROLASES. THEY ARE UNIQUE IN THAT THEY CAN HYDROLYSE
CC CRYSTALLINE CELLULOSE IN THE ABSENCE OF ENDOLUCANASES.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL
CC HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M16190; AAA34210.1; -;
DR EMBL; M55080; AAA72922.1; -;
DR PIR; A26472; A26472.
DR PIR; A26160; A26160.
DR PDB; 3CBH; 15-JAN-91.
DR PDB; 1CB2; 19-MAR-99.
DR Interpro: IPR000254; CBD_fungal.
DR Interpro: IPR001524; Glyco_hydro_6.
DR Pfam; PF00734; CBD_1; 1.
DR Pfam; PF01341; Glyco_hydro_6; 1.
DR PRINTS; PR00733; GLYHYDRLASE6.
DR PRODOM; PD001821; CBD_fungal; 1.
DR PRODOM; PD003733; Glyco_hydro_6; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
KM Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal;
KM 3D-structure. 1
FT SIGNAL 25 471
FT CHAIN 25 471
FT DOMAIN 25 65
FT DOMAIN 66 106
FT DOMAIN 107 471
FT ACT_SITE 199 199
FT ACT_SITE 245 245
FT ACT_SITE 425 425
FT DISULFID 34 51
FT DISULFID 45 61
FT DISULFID 200 259
FT DISULFID 392 439
FT CARBOHYD 38 38
FT CARBOHYD 111 111
FT CARBOHYD 121 121
FT CARBOHYD 130 130
FT CARBOHYD 133 133
FT CARBOHYD 134 134
FT CARBOHYD 139 139
FT CARBOHYD 313 313
FT CARBOHYD 334 334
FT CARBOHYD 199 199
FT MUTAGEN 245 245
FT MUTAGEN 359 359
FT CONFLICT 449 449
FT CONFLICT 449 449
SQ SEQUENCE 471 AA; 49653 MW; C4711BC35B1BD88 CRC64;
Query Match 8.48; Score 548; DB 1; Length 471;
Best Local Similarity 31.28; Pred. No. 1; De-15;
Matches 156; Conservative 60; Mismatches 176; Indels 108; Gaps 17;
QY 7 SGRNCRYQ-----GTRMPAIRSLRGLAGVAVSIASIVP-----LAMQ 48
DB 47 SGRICVSNNDYISQCLGAASSSSSTRRASTTR-----VSPITSSRSSATPPPGSTTTR 101
QY 49 HPAL--AATHVNDPVAGATFFVNPVWAOEVSANQNTNATLAKKRVSVYSTAVAMD 105

```

Db      102 VPPVSGSRTATYSGNPFVGTTPMANAYVASEVSSLAIPSLTGMATRAAAAVAPSPFMD 161
OY      106 -----RIAINGVNGSGFGLTYYDAAISGQGGTTPVEIVYIDLPRDCAALAS 155
Db      162 TLDKTPLEQOTLADIRTKANKNG--NYAG-----QFVYVDLPDRCAALAS 205
OY      156 NGELPATAAGLQTFEYOTVYDIASILSNPKYSRLIVTIEPDSLPNAVTKNSIQACATA 215
Db      206 NGEYSIADGGVAKY-NVIDTRQIV--VEYSDIRFLVLEPDSLANLTNGTPRCANA 262
OY      216 VPYEGIEVALTKLHAIPVNYIYMDAAHSGWIGPNPNASGYQVQKLNASIGVNGID 275
Db      263 QSAYLECINATATOLN-LPNVAMYLDAGHAGLWGPANDPAQLFANYKKAASPRALR 321
OY      276 GVTYNTANTPLKEPPMTATQOVGPVESANFYQWNPDIADAVADLYSLVAAGFSS 335
Db      332 GLATVNAVYNN-----GWNITSPSPSYTOGNNAVYNEKLYIHALGPLANNGW-S 367
OY      336 SIGMLIDTLRNGMGPNPEPTSTATDVNTFVNOSKIDLRQRIQIMCNGNGAGLQPPQA 395
Db      368 NAFETIDGCRSG-----KQPTG-----QQQWGMWCNVIYGFGRPSA 405
OY      396 SPTDEPNMHLADAYWIKPPGESDGTSAASDPPTGKKSDPMCDPTWTTSYGLTNALPNSP 455
Db      406 NTGD---SLDGSFVWVKGGECDDTSDSAP---RFDSHC-----ALPDALQPPAP 449
OY      456 IAGNFPAPFOGLVANARPA 475
Db      450 QAGAMFOAYFVQLTNNAPS 469

RESULT 9
GUNL_ACICE STANDARD: PRT: 562 AA.
AC PS4583;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Endoglucanase E1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E1)
DE (Cellulase E1) (Endocellulase E1).
OS Acidothermus cellulolyticus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Frankineae; Acidothermaceae; Acidothermus.
OX NCBI_TaxID=28049;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 43068 / 11B;
RA Layton R.A., Himmel M.E., Thomas S.R.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RX X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 42-398.
RX MEDLINE=96346058; PubMed=8718854;
RA Sakon J., Adey W.S., Himmel M.E., Thomas S.R., Kaplun P.A.;
RT "Crystal structure of thermostable family 5 endoglucanase E1 from
RT Acidothermus cellulolyticus in complex with cellobiose."
RT Biochemistry 35:10648-10660(1996)
CC -1- FUNCTION: THERMOSTABLE ENZYME WITH AN OPTIMAL TEMPERATURE OF 81
CC DEGREES CELSIUS. HAS A VERY HIGH SPECIFIC ACTIVITY ON
CC CARBOXYMETHYLCELULOSE.
CC -1- CATALYTIC ACTIVITY: Endohydrolasis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL contribution -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR      EMBL: U33212; AAA75A77.1; -
DR      PDB: 1ECE; 14-OCT-96; CBD-2.
DR      InterPro: IPR001919; CBD_2.
DR      InterPro: IPR001547; Glyco_hydro_f5.
DR      Pfam: PF00553; CBD_2; 1.
DR      Pfam: PF00150; cellulase; 1.
DR      PROSITE: PS00659; GLYCOSYL_HYDROL_F5_1.
KW      Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.
FT      SIGNAL 1 41
FT      CHAIN 42 562
FT      DOMAIN 42 400 ENDOGLUCANASE E1.
FT      DOMAIN 401 461 CATALYTIC.
FT      DOMAIN 462 562 PRO/SER/THR-RICH (LINKER).
FT      ACT_SITE 203 203 CELLULOSE-BINDING (BY SIMILARITY).
FT      ACT_SITE 323 323 PROTON DONOR.
FT      DISULFID 75 161 NUCLEOPHILE.
FT      DISULFID 209 212
SQ      SEQUENCE 562 AA; 60747 MW; 84E6256406A35041 CRC64;

Query Match 8.3%; Score 543; DB 1; Length 562;
Best Local Similarity 35.4%; Pred. No. 3e-15;
Matches 138; Conservative 48; Mismatches 104; Indels 100; Gaps 16;

OY      869 GVTYVOGDERTYRQTNEM-----NSSAQOCLTINTAT-----GAWYVSTANFSGTGAPAT 919
Db      243 GVQSTNGDSY-----WMGNLQAGQYPPVILNPNRLVYSADYATSYTP-QTWFSQPT 295
OY      920 YPSIYKGC---HMGCTTKVNGMPIQISGAVTSWSTTVSSGAYDAVDIMNSTPT 976
Db      296 FPNMNPGLINKKWKGLFNNINL-PWILGEGTILDS----- 330
OY      977 TTGQPNGETEIMILNLSRGVOPFGSOTATGTVAGHTMNVWOGQOTSMKIISVYLPGAT 1036
Db      331 TYDQ-----TWLKITL--VQYLRPTAQYAD--SFQWTFW-----SWN----- 363
OY      1037 SISNDLKAIFDAARGLNLSND-----YLDVEAFETIMOGGGLGNSFSVS 1086
Db      364 -----PDSGDTGILKDDMQYVDTVKDGILAPLKSSIFDPVGSASPSQSPSS 412
OY      1087 VT-----SGTSSPTSPSPPTPPSPPTP--TPSPSPPTSPSPSP--SSSGVACRATYYVN 1138
Db      413 VSPSPSPSSASRTPTPTPTPTASPTPTLTPTATPTPTASPTSPSPSPASGACTASTOVN 472
OY      1139 SDWSGFTATVTVTNTGSRATNGVTAVMSFGNQTVTNTWNTALTOGSAATATNLSTNN 1198
Db      473 SDWNGFTVTVAVTNSGVAATKTWTSWTFGNGQTITNSNAAYVQNGSVYARRNNSYNN 532
OY      1199 VTPOGQSTFGFNGSVSGTNAAPTLSCTAS 1228
Db      533 VTPOGNTTFGFQASTYGSNAAPTVACAAS 562

RESULT 10
GUNB_CALSA STANDARD: PRT: 1039 AA.
AC P10474;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endoglucanase/exoglucanase B precursor (includes: Endoglucanase
DE (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase)
DE (Cellobiohydrolase); Exoglucanase (EC 3.2.1.91) (Exocellobiohydrolase)
DE (1,4-beta-cellobiohydrolase)).
DE CELB.
OS Caldicellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Bacilli; Clostridium group;
OC Thermoclostracter group; Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89098398; PubMed=2789517;
RA Saul D.J., Williams L.C., Love D.R., Chamley I.W., Bergquist P.I.;

```

```

RT *Nucleotide sequence of a gene from Caldocellum saccharolyticum
RT encoding for exocellulase and endocellulase activity.*
RL Nucleic Acids Res. 17:439-439(1989).
CC -1- FUNCTION: THIS PROTEIN IS MADE UP OF TWO DOMAINS: THE N-TERMINAL
CC DOMAIN HAS EXOGLUCANASE ACTIVITY WHILE THE C-TERMINAL DOMAIN IS
CC AN ENDOGLUCANASE.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellobiose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO CELLULOSE FAMILY
CC F (FAMILY 10 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION, BELONGS TO CELLULOSE FAMILY
CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X13602; CAA31936.1; -.
DR PIR: S02711; S02711.
DR HSSP: Q06851; INBC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001000; Glyco_hydro_10.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF00331; Glyco_hydro_10; 1.
DR PRINTS: PR00134; GLYDRLASE10.
DR ProDom: PD001947; CBD_3; 1.
DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
DR Cellulose degradation; Hydrolyase; Glycosidase; Repeat;
KW Multifunctional enzyme; Signal.
FT SIGNAL 1 28
FT CHAIN 29 1039
FT DOMAIN 376 416 THR/PRO-RICH, TANDER REPEATS OF T-P.
FT DOMAIN 417 570 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 571 618 THR/PRO-RICH, TANDER REPEATS OF T-P.
FT ACT_SITE 177 177 PROTON DONOR (POTENTIAL).
FT ACT_SITE 285 285 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 792 792 BY SIMILARITY.
SQ SEQUENCE 1039 AA; 117641 MW; 0E0378171594DAB CRC64;

```

```

Query Match 8.0%; Score 519.5; DB 1; Length 1039;
Best local similarity 23.6%; Pred. No. 4,6e-14;
Matches 133; Conservative 98; Mismatches 179; Indels 219; Gaps 21;

```

```

OY 832 NTSAPSPVATATTTSPSPPTGTVDCTPGFNGVTSVQDEXVOTNEMSSAQO 891
DB 597 -----PTAPQTALPPTLPNPIPTSSIPDD-----INDMLYSGN 635
OY 892 CLTINTATGAMVSTANFSGCTGAPATPSIYKGCWMCCTKN-----VGM 939
DB 636 KIVDKDGRPWLGINNFVNTG-----INVEGCV-W-SCNLLDTAEIANGFNLLRV 687
OY 940 PDIISQASVATSMSTQVSSGAV---DVAYDIKNTSPPTTGPON----- 982
DB 688 PIS---AEILINW-----SGGTPKPNINYY---NPELEGKNSLVEDIYVOTKEV 734
OY 983 GTEIMWLNRSRGVQPGSOTATGVTAGCTWVWOCQOTS----- 1023
DB 735 GLKIMLDIHS-----IKTDAMGHILPYWDEKPTPEDFYKACEMITRYKND 782
OY 1024 -----WKIISVYLPATSTISNLDLKAIFDAAARGLSNTSDVLLDVAG 1068
DB 783 TIIAFDLKNEPHGKPMQDTTFKWDNSTDINNMKYA---AETCAKRLININPILLIIEG 839
OY 1069 FE-----IMGGGGLGSNSFSVSTSGTSPFPSP 1098
DB 840 IEAVPKDDVTWTSKSSDYSTWVGWNLGVRKYPINLGRYKQNVYSP 888

RESULT 11
ID GUND_CELFI STANDARD; PRT: 747 AA.
AC P50400;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (cellulase).
GN CEND.
OS Cellulomonas faml.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococccineae; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1708;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93209933; PubMed=8458833;
RA Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C., Jr., Warren R.A.;
RT "Cellulose-binding polysaccharides from Cellulomonas faml: endoglucanase
RT D (Cend), a family A beta-1,4-glucanase."
RU J. Bacteriol. 175:1910-1918(1993).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- PATHWAY: CELLULOSE DEGRADATION.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L02544; AAA23089.1; -.
DR HSSP: P07986; IEXG.
DR InterPro: IPR001919; CBD_2.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003962; FnIII_repeat.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR Pfam: PF00553; CBD_2; 1.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF00410; fn3; 2.

```

DE Mucin 2 precursor (Intestinal mucin 2).
CN MUC2 OR SMOC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R., Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor.";
RT J. Biol. Chem. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=93016075; PubMed=1400449;
RA Gum J.R., Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA Kim Y.S.;
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region.";
RT J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX MEDLINE=91358717; PubMed=1885763;
RA Toribara N.W., Gum J.R., Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA Petersen G.M., Kim Y.S.;
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT and polymorphism.";
RL J. Clin. Invest. 88:1005-1013(1991).
CC -1- PUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
CC OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A
CC PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS
CC AGENTS AT MUCOSAL SURFACES.
CC -1- SUBUNIT: MULTIMERIC.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC BRONCHUS, CERVIX AND GALL BLADDER.
CC -1- PWM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC -1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
CC VARIES AMONG DIFFERENT ALLELES.
CC -1- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT
CC OF SILKORM HEMOCYTIN.
CC -1- SIMILARITY: CONTAINS 2 WFC DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC -----
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL: L21998; AAB95295.1; -;
DR EMBL: M74027; AAA59875.1; -;
DR EMBL: M94131; AAA59163.1; -;
DR EMBL: M94132; AAA59164.1; -;
DR MIM: 158370; -;
DR InterPro: IPR000359; Cys_knot.
DR InterPro: IPR000561; GGF-like.
DR InterPro: IPR002400; GF_cysknot.
DR InterPro: IPR001007; WFC.
DR InterPro: IPR001846; Wvd.
DR Pfam: PF00007; Cys_knot; 1.
DR Pfam: PF00094; Wvd; 4.
DR PRINTS: PR00438; GFCYSKNOT.
DR SMART: SM00214; WVC; 2.
DR SMART: SM00216; WVD; 4.
DR PROSITE: PS00022; GGF_1; UNKNOWN_1

DR PROSITE: PS01185; CTCK_1; 1.
 DR PROSITE: PS01225; CTCK_2; 1.
 DR PROSITE: PS01208; VMFC; 2.
 KM Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 5179
 FT DOMAIN 1401 1747
 FT REPEAT 1401 1416
 FT REPEAT 1417 1432
 FT REPEAT 1433 1448
 FT REPEAT 1449 1464
 FT REPEAT 1465 1471
 FT REPEAT 1472 1478
 FT REPEAT 1479 1494
 FT REPEAT 1495 1517
 FT REPEAT 1518 1533
 FT REPEAT 1534 1556
 FT REPEAT 1557 1572
 FT REPEAT 1573 1596
 FT REPEAT 1597 1612
 FT REPEAT 1613 1635
 FT REPEAT 1636 1651
 FT REPEAT 1652 1675
 FT REPEAT 1676 1683
 FT REPEAT 1684 1699
 FT REPEAT 1700 1715
 FT REPEAT 1716 1731
 FT REPEAT 1732 1747
 FT DOMAIN 4815 4886
 FT DOMAIN 4924 4991
 FT DOMAIN 5075 5160
 FT DISULFID 5075 5122
 FT DISULFID 5089 5136
 FT DISULFID 5098 5152
 FT DISULFID 5102 5154
 FT DISULFID 5159
 FT CARBOHYD 163 163
 FT CARBOHYD 423 423
 FT CARBOHYD 670 670
 FT CARBOHYD 770 770
 FT CARBOHYD 894 894
 FT CARBOHYD 1139 1139
 FT CARBOHYD 1154 1154
 FT CARBOHYD 1215 1215
 FT CARBOHYD 1230 1230
 FT CARBOHYD 1246 1246
 FT CARBOHYD 1787 1787
 FT CARBOHYD 1820 1820
 FT CARBOHYD 4339 4339
 FT CARBOHYD 4351 4351
 FT CARBOHYD 4362 4362
 FT CARBOHYD 4373 4373
 FT CARBOHYD 4422 4422
 FT CARBOHYD 4438 4438
 FT CARBOHYD 4502 4502
 FT CARBOHYD 4616 4616
 FT CARBOHYD 4627 4627
 FT CARBOHYD 4752 4752
 FT CARBOHYD 4787 4787
 FT CARBOHYD 4881 4881
 FT CARBOHYD 4888 4888
 FT CARBOHYD 4955 4955
 FT CARBOHYD 4970 4970
 FT CARBOHYD 5019 5019
 FT CARBOHYD 5038 5038
 FT CARBOHYD 5069 5069
 FT CARBOHYD 1351 1351
 FT CARBOHYD 1412 1412
 FT CARBOHYD 1449 1449
 FT CARBOHYD 1504 1504
 FT CARBOHYD 4192 4192
 FT CARBOHYD 5179 AA: 540295 MW: 85CDJ571FB9A5663 CRC64:

Query Match 7.6%; Score 493.5; DB 1; Length 5179;
 Best Local Similarity 22.6%; Pred. No. 2,36-12;
 Matches 219; Conservative 114; Mismatches 373; Indels 265; Gaps 33;
 QY 392 PPOASPDPEFNAHLDAYVMIKPPGES--DGTSAASDPTTKKSDPMCDPRTYSYGLV 448
 DB 1502 PPMHTT-----IIPPASTTLPLPTTSPPTTTTTPP---PTTSPSPPTT 1545
 QY 449 NALNSPIAGQWPAQFQVLVANAARAVPTSSSPR-PPSPSASRSP-----SP 501
 DB 1546 PTP-----PTSTTLPLPTTSPPTTTPPTTTPPTTTPSPPTTTPPTTTP 1596
 QY 502 SPSSSPSP-----SPSPSSSPS-----PSPPSPSPSSSPSSPSS 538
 DB 1597 PPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1656
 QY 539 PSP-----SPSP-----SPSPSSSPS-----PSBSSSPS-----PS 564
 DB 1657 PSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1716
 QY 565 ----PSPPSPSSSP 619
 DB 1717 STTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1776
 QY 620 LST---VTFRY-----FTRDGSSTLYTN-C--DMAAGCGNIRASFGSVNPT 663
 DB 1777 TPTTPCPVPLCMTWGMDSGRKFNFKPGDTELIGDVGCPGMAA---NISCR-ATMYPDV 1831
 QY 664 PTADTYLQSLFTGGLAAGSGTGEIONRKNSDMSNFETINDSYGNT---TFQDWT 718
 DB 1832 P-----IGLQGVTVWCVDSVGLICKNEQKPGGVIMARCLANEINQCECVTQPTT 1884
 QY 719 KVVYVNVKVLWGEPSSTSPSPSPSP-----SPSPSGGVTPPS 761
 DB 1885 MTPTTTPNPTPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1944
 QY 762 VPTGLVTVGSGSSSLAMNASTDVGVAHNVANGVLVGOPTVSTTDGLAAGT--- 818
 DB 1945 TPITTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1988
 QY 819 ----AYTVAAVDAAGNTSAPSTPVATTTSPSPSPPTPG-----TVVDCATPG 864
 DB 1989 TTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 2048
 QY 865 PNQGVISVQDEVRVQVQNMENSSAQOCLTNTATGATVSTANSGTGCAPIPSIT 924
 DB 2049 PTPPTG-----QTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 2098
 QY 925 KGHMGNCTTKNMGAPIQISQISGAVTSMSTVOSSGAVDAVDIWNSTPTTGOPTGT 984
 DB 2099 -GTQPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 2152
 QY 985 EIMTWLNSRGVQPGSGQTATGVAVGHTWNVWOGQOTSMKLIISVLTGATISINDLK 1044
 DB 2153 PTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 2189
 QY 1045 AIFDAAARGSLNSTDVLDVAGFEIWQGGGGLGASNSFSVSGSSPPSPSPPTTP 1104
 DB 2190 -----TGQTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 2218
 QY 1105 SPTP-----TPSPPTSPSPPTSPSPSGVACATVAVVSDMGSGTAVTVTNGSR 1157
 DB 2219 TTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 2270
 QY 1158 ATNGWVAWSFGNQVTVNNTALVTOGASVATNLSTNNVNIQPGOSTTGFGNSYSGT 1217
 DB 2271 TTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 2330
 QY 1218 NAAPTLSCTAS 1228
 DB 2331 QTPPTTTPPTT 2341

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:21:50 ; Search time 125.36 Seconds

(without alignments)
1694.622 Million cell updates/sec

Title: US-09-917-384-1

Perfect score: 6525
Sequence: 1 MERTQSGSRNCRYGRGTTM.....GFNGSYSGTNAATLCTAS 1228

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seags, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SP archaea:*
- 2: SP bacteria:*
- 3: SP fungi:*
- 4: SP human:*
- 5: SP invertebrate:*
- 6: SP mammal:*
- 7: SP mhc:*
- 8: SP organelle:*
- 9: SP phage:*
- 10: SP plant:*
- 11: SP rodent:*
- 12: SP virus:*
- 13: SP vertebrate:*
- 14: SP unclassified:*
- 15: SP rvirus:*
- 16: SP bacteriap:*
- 17: SP archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1358	20.8	683	16	09pdm2 xyella fas
2	1051	16.1	579	2	086730 streptomyc
3	1015.5	15.6	454	2	09kh72 thermomonos
4	1000	15.3	596	2	060029 thermomonos
5	743.5	11.4	1711	2	P96311 anaerocellu
6	737.5	11.3	930	2	09RFX5 caldicellu
7	722.5	11.1	351	2	09S3V4 cellulomona
8	708.5	10.9	1751	2	09AGC4 caldicellu
9	705	10.8	490	3	096V97 orpiniomyc
10	694	10.4	491	3	096V98 orpiniomyc
11	679	10.3	921	2	091818 caldicellu
12	675	10.3	1779	2	052374 caldicellu
13	669.5	10.3	381	2	09RKY3 streptomyc
14	668.5	10.2	381	2	054331 streptomyc
15	663.5	10.2	1426	2	09X3P6 caldicellu
16	654	10.0	377	2	008468 streptomyc

17	653	10.0	1770	2	09X3P5 caldicellu
18	649.5	10.0	382	2	059663 streptomyc
19	618.5	9.5	460	3	002321 phanerocha
20	618	9.5	371	2	09K1H1 streptomyc
21	597.5	9.2	444	3	096VU2 lentinula e
22	590	9.0	1000	2	024820 thermophil
23	589.5	9.0	443	3	09C1R4 lentinula e
24	582	8.9	436	3	09P8N1 coriolus ve
25	569	8.7	465	3	096RP4 pleurotus s
26	565.5	8.7	476	3	09C1S9 humicola in
27	556.5	8.5	384	2	09X602 streptomyc
28	554	8.5	471	3	09HEX8 trichoderma
29	550	8.4	2332	5	P91365 caenorhabd
30	542	8.3	956	2	09AQH0 caldicellu
31	540	8.3	457	3	093837 acromonium
32	537	8.2	472	3	093860 piromyces r
33	510.5	7.8	1915	2	09RPL0 acetivibrio
34	492.5	7.5	170	2	09RXX6 caldicellu
35	473.5	7.3	473	10	039620 chlamydomon
36	452	6.9	458	3	09UW11 piromyces r
37	444.5	6.8	449	3	P78721 orpiniomyc
38	443	6.8	460	3	09UW10 piromyces r
39	436	6.7	428	3	012646 neocallimas
40	431.5	6.6	997	2	09Z4I1 bacillus sp
41	431.5	6.6	2316	2	09FDP9 bacteroides
42	419	6.4	260	2	033897 rhodotherm
43	417	6.4	376	3	09P808 piromyces r
44	415.5	6.4	261	2	09AGC7 caldicellu
45	412	6.3	459	3	P78720 orpiniomyc

ALIGNMENTS

RESULT ID	1	PRELIMINARY:	PRT:	683 AA.
09pdm2				
AC	09pdm2	01-OCT-2000 (Tremblrel. 15, Created)		
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)			
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)			
DE	1.4-BETA-CELLULOBIOSIDASE.			
GN	XF1267.			
OS	Xyella fastidiosa.			
OC	Bacteria: Proteobacteria; gamma subdivision; Xanthomonas group;			
OX	Xyella			
RN	NCBI_TaxID=2371;			
RP	[1]			
RC	SEQUENCE FROM N.A.			
RX	STRAIN=9A5C;			
RA	MEDLINE=20365717; PubMed=10910347;			
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,			
RA	Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,			
RA	Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,			
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carreiro D.M., Carter H.,			
RA	Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,			
RA	Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Porty H.,			
RA	Fachinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,			
RA	Fraga J.S., Franca S.C., Franco M.C., Frome M., Furian L.R.,			
RA	Gardner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,			
RA	Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,			
RA	Kriegler J.E., Kuramae E.E., Lalget F., Lambais M.R., Leite L.C.C.,			
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,			
RA	Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,			
RA	Martins M.V., Martins E.A.L., Martins E.M.F., Matsushima A.T.,			
RA	Menck C.F.M., Miracca E.C., Miyaki C.Y., Montello-Vilarello C.B.,			
RA	Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,			
RA	Nhani A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,			
RA	de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,			
RA	Pelicko B.R., Pereira G.A.G., Pereira H.A.Jr., Resguero J.B.,			
RA	Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,			
RA	de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,			
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,			

RA de Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terezi M.F., Trufi D., Tsai S.M., Tsunako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Secubal J.C.,
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"
 RL Nature 406:151-159(2000).
 DR EMBL: AE003960; AAF84076.1; -
 DR HSP; P07987; ICB2.
 DR InterPro: IPR001524; Glyco_hydro_6.
 DR Pfam: PF01341; Glyco_hydro_6; 1.
 DR PRINTS: PR00733; GLHYDLASE6.
 DR ProDom: PD003733; Glyco_hydro_6; 2.
 KM Complete proteome.
 SO SEQUENCE 683 AA; 70852 MW; DC8FB5F76B8E8D5F CRC64;

Query Match 20.8%; Score 1358; DB 16; Length 683;
 Best Local Similarity 43.9%; Pred. No. 2,6e-57;
 Matches 319; Conservative 81; Mismatches 218; Indels 108; Gaps 19;

QY 53 AATHVDNRYAGATFFVNPYMAOEVSEANQNTATLAKMRVSTYSTAVMMRIAING 112
 D 18 AEAHVDNPFVATSTYLNPDYSKDIDTSIAKYNDVSLAKEMQITKYPGVWLDITDAVKG 77
 QY 113 VNGGPG--LATTYDAALSOOGTPEVIEIYVDLPERDCAALASNGELPATAAGLQTY 169
 D 78 GPKNDGRNLGHDALAAOKNGKNTPATYATYIDIPGRDCHALASNGELPTEGDLRY 137
 QY 170 ETQYIDPIASILSNPKYSLRIIVTIEPDSLPAVATNMSIOACATA--VPYEGGIEYAL 227
 D 138 KREYIDTASIFSNNKYKDIRLVNIEPDSLPIITNMTSPSCAOKAKMTGIEBEIKTAL 197
 QY 228 TKLAIPNVYIYMDAASGMLGMPNNAAGYVOEOKVL--NASIGVNGIDGFTVTANTTP 286
 D 198 NKLSIPIVNYNMDIGHSGMLGMDNRIPAVSLYTKVIOSTAGPASYNGFATNTANTTP 257
 QY 287 LKEPPM--TATQOVGGQPVSEANFYQWNPDIADAVDLYSLYLAAGFPSSIGMLIDTLR 345
 D 258 LIERPLPMDLNIQGPTRSSKFTYEMNRYPEMDISELLYNDYFAAGPSSIGFTIDYGR 317
 QY 346 NGMGGPNPTPTATDVNTEFVNSKIDLRQHRGLMCNONGAGLGOPPOASPTDEPNALH 405
 D 318 NGMGGPERPTS--AFGNDVNSYVNSGRIDRRHNRGMCNQKAGIGLPIAT----PGGHV 372
 QY 406 DAYVIRKPPGSDGTSASASDPTTGKSDPCDPTTYSYGYLTNALPNSPIAGOWFPNAQF 465
 D 373 DAFWIKRPPGYSDGSSSLIPNDQKGFDRYCDPTFTPDGYLTGALPAPLISGDFHQAQF 432
 QY 466 DOLVANARPAVPTSTSSPPPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 525
 D 433 VKLINNAIPDI-----SRTEPLPGSGSSSSSGSSSSSGSSSSSGSSSSSGSSSSSS 486
 QY 526 SPSSSP 585
 D 487 SGSSSSSSSSSGSSSSSGSSSSSGSSSSSGSSSSSGSSSSSGSSSSSGSSSSSGSSSS 546
 QY 586 GGLKQYQYNNDSAPBDNLIKPLQLVNTGSSSVLDISTYVTRKWFTRDGGSSSTLYVNCMA 645
 D 547 SG-----SSSSSGSSS-----SSGSSS-----SSSGSSGN----- 572
 QY 646 AMGCG--NIRAFSGSVNPTPATDYTLQLSFTGTLAAGSGTEIGIONRKNKMSNFEETN 704
 D 573 -PGAGFNGAGF--NP-----GAGPNPAGSGPTKNPIT-----SSITVDS 610
 QY 705 DYSYG-----TNTTFQDWTKVTYVNGVL--VMGT----- 732
 D 611 DMHTGTCERVKYNTNGSSRSRWSWT--VTIDPKGTIGQTLMSATMSLSSDKLIASGLDMNTL 669
 QY 733 ERSSTS 738
 D 670 EPNGT 675

RESULT 2
 ID 086730 PRELIMINARY; PRT; 579 AA.
 AC 086730;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PUTATIVE SECRETED CELLULOSE.
 GN SC5C7.33.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID-1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE-97000351; PubMed-8843436;
 RA Redenbach M., Kleiser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL; AL031515; CAA20645.1; -
 DR HSP; P07986; IEXG.
 DR InterPro: IPR001524; Glyco_hydro_6.
 DR InterPro: IPR001919; CBD_2.
 DR Pfam; PF00553; CBD_2; 1.
 DR Pfam; PF01341; Glyco_hydro_6; 1.
 DR PRINTS; PR00733; GLHYDLASE6.
 DR ProDom; PD003733; Glyco_hydro_6; 2.
 DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
 SO SEQUENCE 579 AA; 60831 MW; 5A51DD28E7996A89 CRC64;

Query Match 16.1%; Score 1051; DB 2; Length 579;
 Best Local Similarity 49.2%; Pred. No. 9.5e-43;
 Matches 215; Conservative 61; Mismatches 131; Indels 30; Gaps 8;

QY 50 PAIATHVDNRYAGATFFVNPYMAOEVSEANQNTATLAKMRVSTYSTAVMMRIAING 109
 D 158 PTGPEGRDNDNPDGQAVYVNPENMSANAAEPGDR-----IADEPTGVWLDRIA 207
 QY 110 INGVNGSGGLTYYDAALSOOGTPEVIEIYVDLPERDCAALASNGELPATAAGLQTY 169
 D 208 IEGANGGGLDHDAL--EQKSGSEMYQVLYTLPRDCSALASNGELPTE--IDRY 264
 QY 170 ETQYIDPIASILSNPKYSLRIIVTIEPDSLPAVATNMSIOACAT-----APYEGG 222
 D 265 KTEYIDPIAELISDSKVDLRIVTVEIDSLPNIYVNSGTPATENDVKNANGNYOKG 324
 QY 223 IEYALTKHAIPNVYIYMDAASGMLGMPNNAAGYVOEOKVLNASTG--NNGIDGFTYN 280
 D 325 VGIALNRLGAAGVNYNYDAGHGLGWDN--FGPSAIFPTAATTEGATLDVHGFTVN 383
 QY 281 TANYPLKEPWTATQOVGGQPVESANFYQWNPDIADAVDLYSRVAAFPSSIGML 340
 D 384 TANYALAKEENKRTIDSVNGSVRSQSDWYDNNQYTDDELSTYQAMDKLYSLGFDQNLGML 443
 QY 341 IDTLRNGWGNPTPTATDVNTEFVNSKIDLRQHRGLMCNONGAGLGOPPOASPTDF 400
 D 444 IDTSRNGWGNADRPYGPATPTDVNTYVNGGRYDRIRHIGNMCNONGAGLGERPQASPA-- 501
 QY 401 PNAHIDAYVIRKPPGSDGTSASASDPTTGKSDPCDPTT--TSYGYLTNALPNSPIA 457

```

Db 502 --AGIDAYVWMPGSESSSSKLNPDGKGFDRMCDPPTTGNERNGNSMGALPDAPIS 559
Oy 458 GOMFPAQFDOLVANARP 474
Db 560 GAMFSAQFOELMKNATP 576

```

```

RESULT 3
OyK72 PRELIMINARY; PRT: 454 AA.
ID 09K72:
AC 09K72: 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELLULASE PRECURSOR.
OS Thermomonospora fusca.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Thermomonosporaceae;
OC Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RA Al Y.-C., Wilson D.B.;
RT "Genomic DNA sequence encoding Thermomonospora fusca cellulase with
overexpression."
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF268074; AAF75786.1; -.
DR HSSP; P07987; ICB2.
DR Interpro; IPR001524; Glyco_hydro.6.
DR Pfam; PF01341; Glyco_hydro.6; 1.
DR PRINTS; PR00733; GLHYDRASE6.
DR Prodom; PD003733; Glyco_hydro.6; 2.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
KW Signal.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 454 CELLULASE.
SQ SEQUENCE 454 AA; 48624 MW; F8CD1C4CD4807134 CRC64;

```

Query Match 15.6%; Score 1015.5; DB 2; Length 454;
 Best Local Similarity 47.3%; Pred. No. 3.6e-41;
 Matches 221; Conservative 60; Mismatches 147; Indels 39; Gaps 12;

```

Oy 27 LKAGVLAVGAVSTIASTIVPLAMQHPAIAAH-----VDNPAAGAFEFNPNYMAQEVQSEAN 82
Db 6 LKA--LKGAA--AAVLSAALAPPSQAAAPGKVDNPEGAKLYNPVMSAKAAAPGCG 62
Oy 83 QTNATLAAKRVVSTYSTAVMMDRIAAING-----VNGGPGLTYYDLDAALSOOGTTPETV 138
Db 63 S-----AVANESTAVWMDRIGALIEGNSPTGSGMDHLEAV--ROSGGPRLT 111
Oy 139 EIVYIDLPERDCAALASNGELPATAGLQTYETQYIDPIASILSN--PKYSSLRIVYIIEP 197
Db 112 QVYVILPFRDCAALASNGELGPDE--LDRIKSEYIDPIADIMDPADENLRIVAIIEI 169
Oy 198 DSLPNAVMTMSIQACATANPY-----YEGITEYATTKLHAIPNVYIYMDAASHGLGPR 251
Db 170 DSLPNAVMTMSIQACATANPY-----YEGITEYATTKLHAIPNVYIYMDAASHGLGPR 229
Oy 252 NNASGYQEVQFVLNAS--IGVNGIDGFTVNTANTYPLKEPFMTATQOVGQPVESANFYQ 310
Db 230 SNFGPSVDIFYEAAANSSTVDYVHGFIANTANTYATVPEYIDVNGTVNGQLRQSKWD 289
Oy 311 WNPDDIDEADYAVDLVSRVLAAGPSSIGMLIDPLRNGMGWGPPEPSPATDVNTFVNOS 370
Db 290 WQVYDELSEFVDLRQALIAKGFRRSDIGMLIDTSRNGMGWGPPEPSPSTDLNTYVDES 349
Oy 371 KIDLROHRLGKMGONAGAGIQQPQASPTDFPNALHDAVYVIRPGSSDGTSAASDPTTGR 430
Db 350 KIDRLRHGKMGONAGAGIGERPTVNP-----PGVDATVWMPGSESSDASSEIRPDECK 405

```

```

Oy 431 KSDPMCDPPTTTSYGVLTN---ALPNSPIAGQWFPQFDOLVANARP 474
Db 406 GFDRMCDPPTTQGNARNGNPNPSGALPNAPISGHWFSAQFRELLANAYP 452

```

```

RESULT 4
ID 060029 PRELIMINARY; PRT: 596 AA.
AC 060029:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE BETA-1,4-EXOCELLULASE PRECURSOR (EC 3.2.1.91).
GN E3.
OS Thermomonospora fusca.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Thermomonosporaceae;
OC Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang S., Lao G., Wilson D.B.;
RT "Characterization of a Thermomonospora fusca exocellulase."
RL Biochemistry 34:0-0(1995).
DR EMBL; U18978; AAA62211.1; -.
DR HSSP; P07986; IEXG.
DR Interpro; IPR001519; CHD_2.
DR Interpro; IPR001524; Glyco_hydro.6.
DR Pfam; PF00553; CBD_2; 1.
DR Pfam; PF01341; Glyco_hydro.6; 1.
DR PRINTS; PR00733; GLHYDRASE6.
DR Prodom; PD003733; Glyco_hydro.6; 2.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
KW Signal; Hydrolase; Glycosidase.
FT SIGNAL 1 38 POTENTIAL.
FT CHAIN 39 596 BETA-1,4-EXOCELLULASE.
SQ SEQUENCE 596 AA; 63547 MW; B0FA5277FE7721E0 CRC64;

```

Query Match 15.3%; Score 1000; DB 2; Length 596;
 Best Local Similarity 47.7%; Pred. No. 2.6e-40;
 Matches 210; Conservative 55; Mismatches 143; Indels 32; Gaps 9;

```

Oy 50 PLIAATHVDPNPAAGAFEFNPNYMAQEVQSEANQTNATLAAKRVVSTYAVMMDRIA 109
Db 172 PTNPGKVDNPEGAKLYNPVMSAKAAAPGCS-----AVANESTAVWMDRIG 221
Oy 110 ING-----VNGGPGLTYYDLDAALSOOGTTPETVEIYIDLPGHDCALASNGELPATAG 165
Db 222 IEGNDSPTGSGMDHLEAV--ROSGGPRLTQVYIYNLPGHDCALASNGELGPDE-- 278
Oy 166 LOTYETQYIDPIASILSN--PKYSSLRIVYIIEPDSLPAVNTKNSIQACATANPY----- 218
Db 279 LDRYKSEYIDPIADIMDPADENLRIVAIIEIDSLPNLVTNVNGGTELCAVYMRQNG 338
Oy 219 YEGITEYATTKLHAIPNVYIYMDAASHGLGPNNASGYVQEQVLNAS--IGVNGIDDP 277
Db 339 YVNGVYALARKLGEIIPNVYIYMDAASHGLGWSNPSVDIFYEAAANSSTVDYVHG 398
Oy 278 VNTANTYPLKEPFMTATQOVGQPVESANFYQWNPDDIDEADYAVDLVSRVLAAGPSSI 337
Db 399 ISNTANTYATVPEYIDVNGTVNGQLRQSKWDMWQVDELSPVDLRQALIAKGFRRSD 458
Oy 338 GMLIDPLRNGMGWGPPEPSPATDVNTFVNOSKIDLROHRLGKMGONAGAGIQQPQASP 397
Db 459 GMLIDTSRNGMGWGPPEPSPSTDLNTYVDESIRDIRHHPGKMGONAGAGIGERPTVNP 518
Oy 398 TDFPNALHDAVYVIRPGSSDGTSAASDPTTGRKSDPMCDPPTTTSYGVLTN---ALPNS 454
Db 519 A-----PGVDATVWMPGSESSDASSEIRPDECKGFDRMCDPPTTQGNARNGNPNPSGALPNA 574
Oy 455 PIAGQWFPQFDOLVANARP 474

```

Db	575	PISGHWFSQPRELLANVP	594
RESULT	5		
ID	P96311	PRELIMINARY;	PRT; 1711 AA.
AC	P96311;		
DT	01-MAY-1997 (TREMBLrel. 03, Created)		
DT	01-MAY-1997 (TREMBLrel. 03, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	ENOGULCANASE A (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE A)		
DE	(CELLULOSE A) (FRAGMENT).		
DE	CELA.		
OS	Anaerocellum thermophilum.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;		
OC	Anaerocellum group; Anaerocellum.		
NCBI	NCBI_TaxID=31899;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-Z-1320.		
RA	Zverlov V.;		
RL	Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.		
CC	-1- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN		
CC	ENOGULCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL		
CC	DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE		
CC	CELLULOSE (BY SIMILARITY).		
CC	-1- CATALYTIC ACTIVITY: ENDOPOLYMEROLYSIS OF 1,4-BETA-D-GLUCOSIDIC		
CC	LINKAGES IN CELLULOSE.		
CC	-1- SIMILARITY: THE N-TERMINAL PART BELONGS TO CELLULOSE FAMILY E		
CC	(FAMILY 9 OF GLYCOSYL HYDROLASES).		
CC	-1- SIMILARITY: THE C-TERMINAL PART BELONGS TO CELLULOSE FAMILY L		
CC	(FAMILY 48 OF GLYCOSYL HYDROLASES).		
DR	EMBL; Z86105; CAB06786.1; -.		
DR	HSSP; P26221; 1TF4.		
DR	InterPro: IPR001956; CBD_3.		
DR	InterPro: IPR000556; Glyco_hydro_48.		
DR	InterPro: IPR001701; Glyco_hydro_9.		
DR	Pfam: PF00942; CBD_3; 3.		
DR	Pfam: PF02011; Glyco_hydro_48; 1.		
DR	Pfam: PF00759; Glyco_hydro_9; 1.		
DR	PRINTS: PR00844; GLHYDRLASE48.		
DR	Prodom: PD001947; CBD_3; 2.		
DR	Prodom: PD011903; Glyco_hydro_48; 1.		
DR	PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.		
DR	PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.		
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.		
FT	NON_TER 1		
SEQ	SEQUENCE 1711 AA; 189979 MW; E3E987CEB9CD0C21 CRC64;		
Query Match	11.4%; Score 743.5; DB 2; Length 1711;		
Best Local Similarity	22.8%; Pred. No. 1.4e-27;		
Matches 277;	Conservative 169; Mismatches 353; Indels 417; Gaps		
35	AVSIASIVPLMAHNPATIAIVNDVPYAGATFFVNPYMAQEVQSEA-ANQTNATYLAAKRR	93	
156	AAASLAASIVLDRNDPHTDAATYLQH-----AKELYERAEVYKSDAGTYAANG	202	
94	VYSTS-----TAVMMDRIAALINGVNGSGLTITTYIDAALSSQGGTPEVEIEIYIDL	145	
203	YNSNSGTFDELSTMAAVWL--YLATND-----STYLTKASIVYO-NMPKISGSNTIDY	252	
146	PGRDAAALASNGE--LPATMAGLQTYETQYIDPIAS-----ILSNPK-----Y	186	
253	KAAHCHMDVDHNAALLAKITTKDKIYK-QIISHLDYITWTGNGERIKITTPGKGLAMLDW	311	
187	SLSLRIYV-----IIEPDSLPAVNTNMSIQACATAVPPYEGOEIYAL-----TKLHA	232	
312	GSILRATYATFAIFYSDVNGCPSTKKEIYR-----KFGESQIDVALSAGRSFYVGEFT	366	
233	INVTYIYMDAAHSGW-----LGMPPNNAISGYDQEVQKVLNLSIGVNCI	274	

[illegible]

DT 01-DEC-2001 (TEMBLrel. 19, last annotation update)
 DE MULTIDOMAIN BETA-1,4-MANNANASE PRECURSOR.
 GN MANA.
 OS Caldicellus celluloovorans.
 OC Bacteria: Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Alicyclobacillus group; Caldicellus.
 ON NCBI_TaxID=74586;
 RX SEQUENCE FROM N.A.
 RP MEDLINE=20120520; PubMed=10653733;
 RA Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
 RT "A gene encoding a novel multidomain beta-1,4-mannanase from
 RT Caldicellus celluloovorans and action of the recombinant enzyme on
 RT Kraft pulp";
 RL Appl. Environ. Microbiol. 66:664-670(2000).
 DR EMBL: AF163837; AAF22274.1; -
 DR HSSP: Q06851; INBC.
 DR InterPro: IPR001956; CBD_3.
 DR InterPro: IPR004302; Chitin_bind_3.
 DR InterPro: IPR001547; Glyco_hydro_F5.
 DR InterPro: IPR002965; P-rich_extensn.
 DR Pfam: PF00942; CBD_3; 2.
 DR Pfam: PF00150; cellulase; 1.
 DR Pfam: PF03067; Chitin_bind_3; 1.
 DR PRINTS: PRO1217; PRICHEXTENS.
 DR PRODOM: PD001947; CBD_3; 2.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 KW signal.
 FT SIGNAL 1 33 POTENTIAL.
 FT CHAIN 34 930 MULTIDOMAIN BETA-1,4-MANNANASE.
 SQ SEQUENCE 930 AA; 101576 MW; 0086638D54D1A2CC CRC64;

Query Match 11.3%; Score 737.5; DB 2; Length 930;
 Best Local Similarity 25.0%; Pred. No. 1.4e-27;
 Matches 246; Conservative 111; Mismatches 270; Indels 357; Gaps 36;

QY 411 IKPESGDSASADPTTGKSDMPDPTT-----SYGLTNALNPSIAGQNF--- 461
 DB 97 IIPGKICGPTASFDGNQARTD-----WMTTRIQPGATITVRNAM--APHGPTWLYV 149
 QY 462 -----PAQFDQLVANRAPVPTSTSSPPPP-----PPSPASPSPPSPSSSSPSP 509
 DB 150 TRDGMDFQ-----PLKMSDLEPTPSQVTPNPINSSGPGCAEYSWQVQLPNK 197
 QY 510 -----SPSP-----SSSPSPSPSPSPSSSPSPSSSPSPSPSPSPSPSPSPSP 546
 DB 198 QGRHIIYMIQKSDSPFAFNCSQVYFGSGPIATFEGDPRBGGTMTTPPSGTTPTPTPT 257
 QY 547 PSPSSSP 606
 DB 258 PTPPTPTPTP-----TPPSPVPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 308
 QY 607 GLQLVNTGSSVDLSTYTVKWTFRDGGSSFLVNCMAAGCGNITASGVSNPATPTA 666
 DB 309 QLRTVNTGSAVPLTELKVTWYTKN--STQAEQYFCMAIGSCNIRAQEVSLSQPVSGA 367
 QY 667 DTYQLSFTGTLAAGSTGEIQRNVKSDMSNFTENDSY--GTNTFDQMTVTVYVN 725
 DB 368 DSTIELSFTGSLPAGNTEIQRNHFTHMNMANTETDMSYNGAQTWGPSTRTILYRN 427
 QY 726 GVLVWGTEPGSTSP 785
 DB 428 GVLVWGTEPGSGSPPT 455
 QY 786 NCVAHNINVRNGVLVQPTVSTDTGLAGTAYTTVAANDAGNTSAPSTVLTATT 845
 DB 456 -----TS 457
 QY 846 SPSPSPPTGTTVTDCTPGPNQNGTVSGDDEVIVQTNEM-----SSAOCCLINTATG 900
 DB 458 TPT 506

QY 901 AMTVSTANFSGGTGAGATPSPSIYKCGHMGNCCTKNMGPIQISQISAVTSMSTGVSS 960
 DB 507 SW-----CANAVRIVLSNGCRM-----TK-----IPASEVADIISOARTLGRA 545
 QY 961 GANDV-----AIDITNSPTPTTGQPNNGEIMWILMSRGVQPPGSGQT 1003
 DB 546 VLEVHDTGTGEDAACSMTTAVNWIETLKNVLAGENFYVNI-----GNPEYGN 599
 QY 1004 ATG-----VTVAGHTVNNVMOGQOSTSWKILSYVLTGATGISNLD- 1042
 DB 600 YQNNVPTFRNAVQALRNAGINTNTINVDAPNN-----GQ--DH--SFTMRDNAPITFNADP 650
 QY 1043 -LKAIFA-----DAAR-----GSLNTS-----DYLDV 1065
 DB 651 QRLNVFSIHMGVYDTAAEYQSTIESFVNGRLPLVIGEFGMHSDGDPNDAQIVQAKQY 710
 QY 1066 EAGFEIWO--GGQIGLGSFSVSYSSTGSSPT-----PSPPTPTPTPTPTPTPTPTPT 1112
 DB 711 NIGLFGWSWGGNGGVEYLDVNTFNANSPFAMGTWERTNAIGTSTPTPTPTPTPTPTPT 770
 QY 1113 SPSPSPSP-----TSPPSSSGVACRATVYV----- 1137
 DB 771 TPTPTPSAGGLVYQYRAADTNATDNLKRFRTVNNCTSSVPLSELTIRKWTYVDGDKP 830
 QY 1138 ---NSDMSGFTATVTVNTGSRATNGMTVYAMSGNQYTVNTWNTALQTSGASVTATNL 1194
 DB 831 QVFCNDM-----AOVGCNLRGSPFKLTSG--RTGADYVIEITFTSGAGSLAPGA 878
 QY 1195 SYNNV-----IQPGOSTTFGFGNGSYS 1215
 DB 879 SSGDIQVRINKNDMTNYNENDYS 902

RESULT 7
 Q9S3V4
 ID Q9S3V4 PRELIMINARY; PRT; 351 AA.
 AC Q9S3V4;
 DT 01-MAY-2000 (TEMBLrel. 13, Created)
 DT 01-MAY-2000 (TEMBLrel. 13, last sequence update)
 DE 01-JUN-2001 (TEMBLrel. 17, last annotation update)
 DE 1,4-BETA-CELLULOBIHYDROLASE A (EC 3.2.1.91) (FRAGMENT).
 GN CELLOMONAS flavigena.
 OS Cellulomonas flavigena.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Micrococccineae; Cellulomonadaceae; Cellulomonas.
 OX NCBI_TaxID=1711;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CD8B531;
 RA Herrera A., Gutierrez A., Saigado L.M., Ponce-Noyola T.;
 RT "Molecular characterization of cellulases from Cellulomonas
 RT flavigena";
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF172344; AAD48493.1; -
 DR HSSP: P07987; ICB2.
 DR InterPro: IPR001524; Glyco_hydro_6.
 DR Pfam: PF01341; Glyco_hydro_6; 1.
 DR PRINTS: PR00733; GLYHDLASE6.
 DR PRODOM: PD003733; Glyco_hydro_6; 1.
 DR PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
 KW Hydrolase; Glycosidase.
 FT NON_TER 351 351
 SQ SEQUENCE 351 AA; 37615 MW; BCD6E3B5DF335C2 CRC64;

Query Match 11.1%; Score 722.5; DB 2; Length 351;
 Best Local Similarity 48.1%; Pred. No. 2.5e-27;
 Matches 166; Conservative 45; Mismatches 105; Indels 29; Gaps 13;

QY 21 PAISKRLRACVLGAVSIASI--VPLAQ--HPLIAA--THVDNRYACATFVNRYNA--QE 75
 DB 6 PSVARRRRTV--GAATTAALVAVPLTLISLSPASAAEARVDNRYAGAVQVYVNTWMASS 63

QY 1099 SPTPTSPPTPTSPSPPTSSPSGACATVYVNSDMGCTATATVNTNGSRA 1158
 ID 096V97 PRELIMINARY; PRT; 490 AA.
 AC 096V97;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CELLOBIOLHYDROLASE II-LIKE CELLULOSE CELI.
 GN CELI.
 OS Orpinomyces sp. PC-2.
 OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
 OC Neocallimastlicaceae; Orpinomyces.
 OX NCBI_TaxID=50059;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PC-2;
 RX MEDLINE=21405733; Pubmed=11514516;
 RA Steenbakkers P.J.M., Li X.L., Ximenes E.A., Arts J.G., Chen H.,
 RA Ljungdahl L.G., Op Den Camp H.J.M.;
 RT "Noncatalytic docking domains of cellulosomes of anaerobic fungi.";
 RL J. Bacteriol. 183:5325-5333(2001).
 DR EMBL; AF177205; AAL0121.1; -;
 KW Hydrolase.
 SQ SEQUENCE 490 AA; 54051 MW; 85FA0D1280759886 CRC64;

Query Match 10.8%; Score 705; DB 3; Length 490;
 Best Local Similarity 39.0%; Pred. No. 2.4e-26;
 Matches 167; Conservative 54; Mismatches 147; Indels 60; Gaps 12;

QY 50 PAATATVNDPVYAGATFVYVNAOYSEANOTNATLAAKRVSTYSTAVMMDRIAA 109
 ID 096V97 PRELIMINARY; PRT; 490 AA.
 AC 096V97;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CELLOBIOLHYDROLASE II-LIKE CELLULOSE CELI.
 GN CELI.
 OS Orpinomyces sp. PC-2.
 OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
 OC Neocallimastlicaceae; Orpinomyces.
 OX NCBI_TaxID=50059;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PC-2;
 RX MEDLINE=21405733; Pubmed=11514516;
 RA Steenbakkers P.J.M., Li X.L., Ximenes E.A., Arts J.G., Chen H.,
 RA Ljungdahl L.G., Op Den Camp H.J.M.;
 RT "Noncatalytic docking domains of cellulosomes of anaerobic fungi.";
 RL J. Bacteriol. 183:5325-5333(2001).
 DR EMBL; AF177205; AAL0121.1; -;
 KW Hydrolase.
 SQ SEQUENCE 490 AA; 54051 MW; 85FA0D1280759886 CRC64;

QY 110 INGVNGSGPLTTYLDALSOOGCTPEVIEI-VIYDLPGRDCAALASNGELPATAGLOT 168
 DB 118 PGEQYTHSGPFGSVEFFLNPYVAEYDAIEONTSSSLAKAKEMKMTYSNAIMLDYIK - 176
 QY 110 INGVNGSGPLTTYLDALSOOGCTPEVIEI-VIYDLPGRDCAALASNGELPATAGLOT 168
 DB 177 -----NMHQMLETNKGALQOGEFTGSKVLVYFVYDLPGRDCHALASNGELANDSDAQR 232
 QY 169 YEQYIDPIASILSNPKYSLRYITIEPDSLPAVNTNM-SIOACATVAPYEEGIEYAL 227
 DB 233 YKTEYIDVIEEKLK--YKSGQPVLLIIEPDSLANTVNTLSTPACRDSSEKYIYLDGHAYLI 290
 QY 228 TKLHAIPVYIYMDAHSGLGMPNNASGYVOEKYVLNASIGVNG-IDGFPVTANTYTP 286
 DB 291 OKLGVLPVAMVYLDIGHAFMLGMDNREKAKYYSKVI--SSGPGVYVKGTDVNAVYTP 348
 QY 287 LKEPFTATQOVGGQPVESANFYQWNPDIDEADYAVDLVSLRYAAGFPSSIGMLIDTLRN 346
 DB 349 WEDEPTLSRGFET-----EWNPCPEDEKRYLELMHKDFKAALIESYEFVCDISRN 397
 QY 347 GCGPNEPTGPSTATDVNTVYVNSKIDLRHGRGLMCNONGAGLGPPOASPTDFPNAHLD 406
 DB 398 G-----HKVD-RKHGEMCNGTGVGVARPOASPVSGMD-YLD 433
 QY 407 AYWIKPGEESDGTSAASDPTTGKKSDDPCDPTTYSYGLVNLALPNSPIAGOWEPAQFD 466
 DB 434 AFWIKPLGESDGT-----DTSARVYDGYCGH-----DTAKMPAPEAGOWFOKHFE 480
 QY 467 QLVANARP 474
 DB 481 GLENANP 488

Q96V98
 ID 096V98 PRELIMINARY; PRT; 491 AA.
 AC 096V98;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CELLOBIOLHYDROLASE II-LIKE CELLULOSE CELI.
 GN CELI.
 OS Orpinomyces sp. PC-2.
 OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
 OC Neocallimastlicaceae; Orpinomyces.
 OX NCBI_TaxID=50059;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PC-2;
 RX MEDLINE=21405733; Pubmed=11514516;
 RA Steenbakkers P.J.M., Li X.L., Ximenes E.A., Arts J.G., Chen H.,
 RA Ljungdahl L.G., Op Den Camp H.J.M.;
 RT "Noncatalytic docking domains of cellulosomes of anaerobic fungi.";
 RL J. Bacteriol. 183:5325-5333(2001).
 DR EMBL; AF177204; AAL0121.1; -;
 KW Hydrolase.
 SQ SEQUENCE 491 AA; 53956 MW; 7C7ACCC9BF946C4 CRC64;

Query Match 10.6%; Score 694; DB 3; Length 491;
 Best Local Similarity 38.6%; Pred. No. 8.2e-26;
 Matches 165; Conservative 53; Mismatches 149; Indels 60; Gaps 12;

QY 50 PAATATVNDPVYAGATFVYVNAOYSEANOTNATLAAKRVSTYSTAVMMDRIAA 109
 DB 121 PGSQTLTHSGPFGSVEFFLNPYVAEYDAIAQMSNSSLAKAKEMKMTYSNAIMLDYIK - 179
 QY 110 INGVNGSGPLTTYLDALSOOGCTPEVIEI-VIYDLPGRDCAALASNGELPATAGLOT 169
 DB 180 -----NMQOMLETNKGALAOQTGSKVLVYFVYDLPGRDCHALASNGELANDSDAQR 235
 QY 170 ETOYIDPIASILSNPKYSLRYITIEPDSLPAVNTNM-SIOACATVAPYEEGIEYAL 228
 DB 236 KTEYIDVIEEKLK--YKSGQPVLLIIEPDSLANTVNTLENTPACRDSQYIYLDGHAYLIK 293
 QY 229 KLAIPVYIYMDAHSGLGMPNNASGYVOEKYVLNASIGVNG-IDGFPVTANTYTP 287
 DB 294 KFGVLPVAMVYLDIGHAFMLGMDNREKAKYYSKVI--SSGSPGKYRGFTDNAVYTPW 351
 QY 288 KEPMTATQOVGGQPVESANFYQWNPDIDEADYAVDLXSLRYAAGFPSSIGMLIDTLRN 347
 DB 352 EDPTLSRGFPT-----EWNPCPEDEKRYLEAMHKDFKAAGI--SSYFVSDTSRNG 399
 QY 348 WCGPNEPTGPSTATDVNTVYVNSKIDLRHGRGLMCNONGAGLGPPOASPTDFPNAHLD 407
 DB 400 -----HKTD-RKHGEMCNGTGVGIGRPPANPISMD-YLDA 435
 QY 408 YWIKPGEESDGTSAASDPTTGKKSDDPCDPTTYSYGLVNLALPNSPIAGOWEPAQFD 467
 DB 436 FYWIKPLGESDGT-----DTSARVYDGYCGH-----ETAKMPAPEAGOWFOKHFE 482
 QY 468 LVANARP 474
 DB 483 GLENANP 489

RESULT 11
 Q96V98
 ID 096V98 PRELIMINARY; PRT; 921 AA.
 AC 096V98;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE BETA-1,4-XYLANASE XYNA PRECURSOR.
 GN XYNA.
 OS Caldicellibacillus cellulovorans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:16:30 ; Search time 47.58 Seconds
(without alignments)
630,404 Million cell updates/sec

Title: US-09-917-384-1
Perfect score: 6525
Sequence: 1 MERTQSGRCRCRYRGRTTRM.....GFNGSYSGTNAPLISCTAS 1228

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/prodata/2/1aa/Backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	708.5	10.9	1751	4	US-09-136-574A-44
2	668.5	10.2	1381	4	US-09-216-295-22
3	663.5	10.2	1426	4	US-09-136-574A-43
4	639.5	9.8	352	4	US-09-286-691-26
5	639.5	9.8	352	4	US-09-687-147-26
6	622.5	9.5	386	4	US-09-321-981-5
7	618	9.5	371	4	US-09-104-308-1
8	618	9.5	371	4	US-09-321-981-1
9	556.5	8.5	461	3	US-09-118-319-9
10	555	8.5	360	4	US-09-286-691-25
11	555	8.5	360	4	US-09-687-147-25
12	552	8.5	312	4	US-09-216-295-21
13	548.5	8.4	470	3	US-09-118-319-8
14	543	8.3	521	1	US-08-276-213-3
15	540	8.3	457	3	US-09-142-759-1
16	538	8.2	360	4	US-09-286-691-24
17	538	8.2	360	4	US-09-687-147-24
18	538	8.2	365	2	US-08-169-948B-12
19	538	8.2	365	2	US-08-448-873-12
20	538	8.2	365	4	US-08-382-452D-12
21	444.5	6.8	449	3	US-09-118-319-7
22	444.5	6.8	449	4	US-09-286-691-4
23	444.5	6.8	449	4	US-09-687-147-4
24	436	6.7	428	3	US-09-118-319-5
25	434.5	6.7	326	4	US-09-286-691-23
26	434.5	6.7	326	4	US-09-687-147-23
27	431.5	6.6	432	3	US-09-118-319-2

28	419	6.4	260	4	US-09-216-295-23	Sequence 23, Appl
29	417.5	6.4	551	2	US-09-033-537A-1	Sequence 1, Appl
30	413.5	6.3	616	4	US-09-136-574A-47	Sequence 47, Appl
31	412	6.3	459	3	US-09-118-319-6	Sequence 6, Appl
32	412	6.3	459	4	US-09-286-691-2	Sequence 2, Appl
33	412	6.3	459	4	US-09-687-147-2	Sequence 2, Appl
34	409	6.3	493	4	US-09-198-956-10	Sequence 10, Appl
35	409	6.3	493	4	US-09-198-955A-12	Sequence 12, Appl
36	381.5	5.8	167	5	PCT-US95-13813-9	Sequence 9, Appl
37	377.5	5.8	700	2	US-07-862-588B-2	Sequence 2, Appl
38	290	4.4	907	3	US-08-783-774-2	Sequence 2, Appl
39	290	4.4	907	5	PCT-US95-04611A-19	Sequence 19, Appl
40	284	4.4	348	4	US-09-216-295-16	Sequence 16, Appl
41	268.5	4.1	1185	4	US-09-041-886-23	Sequence 23, Appl
42	266	4.1	214	1	US-08-217-327-4	Sequence 4, Appl
43	259.5	4.0	259	4	US-09-216-295-5	Sequence 5, Appl
44	255.5	3.9	1719	2	US-08-459-568-4	Sequence 4, Appl
45	255.5	3.9	1719	2	US-08-399-411-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-09-136-574A-44
; Sequence 44, Application US/09136574A
; Patent No. 6294366
GENERAL INFORMATION:
APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.
TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 1751 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 44:

US-09-136-574A-44

Query Match 10.9% Score 708.5; DB 4; Length 1751;
Best Local Similarity 22.3%; Pred. No. 8,7e-37;
Matches 304; Conservative 165; Mismatches 429; Indels 467; Gaps 54;

QY 35 AVSTAASTVPLAMHRAIAATHV-----DNFYAGTFVN----- 69
DB 192 AASTAAVSVVIAKERSOKRAALQAKDLFEADTRSDAGYATATGTYSGGFIIDDLGW 251
QY 70 -PYW-----AOEVOSEANOTNA-----TLAKMRYSTYSTAV 102
DB 252 AAVWLYIATNDSSYLTKAEELMESEVANCNTWTCQMDVRYGTILIMAKITKEIYKAV 311
QY 103 -----WMDRIAINGVNGPGLFTTYDALSQQGTTPEVEIYIYDLPGRDC----- 150
DB 312 ERNLDMHTDRITTYT-----PKMAYILMGWGLSLRATTAALACYYADMGCDSSKTKRY 365
QY 151 -----ALASNGELPATAAGIQYETQYIDPISILNPKXSLRIYITIEPDSL 201
DB 366 LNFASKQIDYALGSTGRSEVVGFG-----TNY-----POH-----P 396
QY 202 NAYTNMSIOACATAVPYEEQIEYALFKIAIPVYIYMDAAHSGMLGMPNNAAGYQOE 261
DB 397 HHRNAHSSMANSMKIPEYHRILY-----GALVCGPGSDSDSNDI 437
QY 262 OKVLNASTIGVNGIDFVNTNTANTYPLKEPMTATQVGQPVESANFYQMPDIDDEADYA 321
DB 438 TDYVONEVACDYNAGIYDALAK-----MYOLYGEPIIDPKALE-TPTNDEI-FV 485
QY 322 VDLVSRILVAAFPSSISGLIDTLRNGMGCPNEPCSTATVNTVNSKIDILAKHRLM 381
DB 486 ESKGNSOGCPYETIYIYIY-RTGW-----PPRYTDLKSL-KYFIDLTSL- 529
QY 382 CNONGAGIGOPPOASPTDFPNAHDA-----YVW-----IKPGESD 418
DB 530 -----IQAGYSPDVYKVD--TYIIEGKISGRPYWDKRNRIYVYLVDSGRIYIGEVE 582
QY 419 GTSASADPTTKKSDPMCDPTTYTTSYGLVNLALPNSPIAGWFPAQFD--OLVANARAV 476
DB 583 HKKQAOEKISVPOGYPM-DPTNDSYKGLTQLEKNK-----YIAAYNNMLVWGLEPGA 636
QY 477 PTSTSSPPPPPSAPSP 536
DB 637 ATST-----PAPT----- 644
QY 537 SSP 595
DB 645 STPTPTPTPTPTATPTPT-----TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 686
QY 596 DSAPGDNOIKFGLQLVNTGSSVDLSTVYRWETTRDGSSTLYVNCMAAMCGCNIRAS 655
DB 687 ETSASTGSIKRWFKIIVNGSSSVDSLIRYKIRMYVVDGDKPOSAV-CMAQAGASNTWFN 745
QY 656 FGSVNPATPTADTYLQISFTGTT--LAAGSTGETIONRVNNSDMSENFETINDYST-GINT 712
DB 746 FVKLSSGVSGADYTLIEVFGSSGAGOLPGKDTIQVRFKNNDMSNYQADMSVLSQMT 805
QY 713 TFODMTKTVVNGVYLVWGTPEPGTSPSP-----TPSPSPSPSPSPSPSPSPSPSP 760
DB 806 NYGEAKAKTLLVVDGLVWVGQERGCATPAFTSTATTPTPTPTPTPTPTPTPTPTPTPT 865
QY 761 SVPTGLVTVGVSGSSSVLSAMNASTDNVGAHYNYRVNGVLVGQPTVSTFTDGLAAGTAY 820
DB 866 PAPT--ASPVGGS--YMPSPS----- 882
QY 821 TTYTAAV--DAAGNTPASPTPTATITSSPSPPTPTGTTVTDCPTSPNONGVTSVQDEXR 879
DB 883 SYGALKVYIANGNLSSTP-----NYLNPKIRIENVGTAVLDS-----R 921
QY 880 VOTNEM--NSSAQQCLINTATGA--WTYSTRANSGGTGGAATYPSIYK--GCHMGN 931

DB 922 VKRYWTTIDGEGATQSVVASSINPATYIDVKFKIGANAGCADYVEIGFKSGACVLAAG 981
QY 932 CTTKNVGMPQIOISQISAVTSMTSTQVSSGAYDAVDIMTSTPT-----TTG----- 979
DB 982 OSTKEIRLSI-----QKSGSYNSQNDVSANSIENKEYTGYIDVL 1026
QY 980 -----QP-NGTEIMTY-----INSRGVOPFSSQATGYTVAGHVMWQGOOT 1022
DB 1027 VWGREGPNAQIKWYANGNIGMTNVLNPKIENVTGTVLDSRVYRVRYTTIDGEGAT 1086
QY 1023 SWKISYLVTPGATISIMLDKAIIPADAARGLSTSYLDV--EAGFEIMQGO----- 1076
DB 1087 OSVATSSINP-----AIDVKFVLGNAGC-----ADYVEIGKSGAGVLAAGOSTKE 1137
QY 1077 -----GLG-----SNSESVTSYG-----TSSPTSP 1098
DB 1138 IRLSIOKSGSYNSQNDVSATSGATIEENKEYTGYIDGAIWGREPSRGTPAGVETPL 1197
QY 1099 SPT 1158
DB 1198 APPT 1242
QY 1159 TNGWTVAMSGGNOTVTYMNMTALTFQSGASYATNLSTNNYIQPG 1203
DB 1243 RTKIS-PTIYGANDIOGVHPARLGNRLTGYNWE--NNMSNAG 1285

RESULT 2
US-09-216-295-22

; Sequence 22, Application US/09216295

; Patent No. 6268328

; GENERAL INFORMATION:

; APPLICANT: Mitchinson, Colin

; TITLE OF INVENTION: No. 6268328el Variant EgIII-like Cellulase Compositions

; FILE REFERENCE: GC555

; CURRENT APPLICATION NUMBER: US/09/216,295

; CURRENT FILING DATE: 1998-12-18

; SOFTWARE: FASTSEQ for Windows Version 3.0

; SEQ ID NO 22

; LENGTH: 381

; TYPE: PRT

; ORGANISM: Streptomyces lividans CelB

US-09-216-295-22

Query Match 10.2% Score 668.5; DB 4; Length 381;
Best Local Similarity 39.1%; Pred. No. 4.1e-35;
Matches 158; Conservative 49; Mismatches 150; Indels 47; Gaps 13;

QY 839 AAGNTSAPSTPYATTTSPSPSPPTPTGTTVTDCTPGPNONGVTSVQDEYRVQTNMNS 888
DB 17 ALGAVLAAPALVSSLVTAAPAOADT-----TICEP-----FGTTTIG-RYVQNNRMGST 67
QY 889 AQQCLITATGAMTVSTANSGGTCGAPATYPSIYKCHCHMNCCTTKVAMPQIQSIGS 948
DB 68 APQCVTA-TDYG-FRYTOADGAPNGAPKPSYFVNCGHYTN-C-SPGTDLPVRIDTVSA 124
QY 949 AVTSMSTQVSSGADVADVDIMTNTPTTGTGPNCTEIMIMVNSGVOGPCSQATGVT 1008
DB 125 APSSISYGFVDCAVYNASDIDMLDPTARTDG-VNQTETIMVNNRVGPQIPGSPVGT-AS 182
QY 1009 VAGHTWVWQGOQTSWKISYLVTPGATISIMLDKAIIPADAARGLSTSYLDV--EAGFEIMQGO 1068
DB 183 VGRFTWVWSSGNGSNDVLSFV-APSAISGWFVDMDVRYATVANGLAENDMYLTSVQAG 241
QY 1069 FEIMQGGGLGNSPVSATSGTSSPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 1128
DB 242 FEPQNGAGLAVNSPSSVETGTGPGT-----DQDPPGGS--- 277
QY 1129 VACRATYVNSWVGSGFTATYVTVTGSRATNGTVAMSGFNGQTVTYMNMTALTFQSGAS 1188


```
Db 342 QSGFGGTSSGAGFTAPACARLNGTSC 369
:|||||:|:|:|
RESULT 9
US-09-118-319-9 : Sequence 9, Application US/09118319
: Patent No. 6114158
: GENERAL INFORMATION:
: APPLICANT: Li, Xin-Liang
: APPLICANT: Chen, Huizhong
: APPLICANT: Ljungdahl, Lars G.
: TITLE OF INVENTION: Orpinomyces Cellulase Celf Protein and Coding Sequences
: FILE REFERENCE: 33-98sequence listing
: CURRENT APPLICATION NUMBER: US/09/118,319
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 9
: LENGTH: 461
: TYPE: PRT
: ORGANISM: Fusarium oxysporum
US-09-118-319-9

Query Match
Best local Similarity 8.5%; Score 556.5; DB 3; Length 461;
Matches 149; Conservative 71; Mismatches 167; Indels 75; Gaps 16;

QY 17 TTRPAPISRLRAGVLAVSIASIVPLMQHPATATHTVDNPYAGATFVNPYMAOEY 76
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 70 TAGPSTSTATKTATGSGSTTAGSVTSN---PPAAS---DNFYAGVDLMANNYRSEV 123
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 77 QSEFA---NOTNATLAKMRVSTYSTAVWMRIAINGVNGSPGLTYYIDAALSOQOQT 133
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 124 MNLAIVKLSGAKATAAKADVPSFO---WMDTYDHSILME---DTLDIRKANRAG- 174
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 134 TPVEIEIYIDLPGRCALASNGELPATAGTQYETQYIDIPASISPKYSRLIYV 193
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 175 GKTAGGVVVDLPNRDCAALASNGEYSLDKGANKRYA-YIAIKIGLON--YSDRKVIL 221
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 194 IIEPDSLPAVNTMSIOACATAVRYEOGIEVALTKLHAIPNYIYMDAHSGLGPNPN 253
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 VIEPDSLANTVTLNVDKCAKASAKELVYAIKELN-LPNVSMILDAHGMLGMPAN 290
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 254 ASGYVEQVKVLAASIGVNGIDCFVTNTANTYPLKEPMTATQOVGQPVESANFYQWNP 313
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 291 IGPAKLYAQIYKDAKPSRVRGLVTNVSNN-----GWLKSTKPDYTESNP 337
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 314 DIDEADYAVDLYSRVLAAGPSSIGMLIDFLRMGNGPNEPTPTATDVNTFVNSKID 373
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 338 NYDEQRY-INAFAFLAQEGSWNVKFLVDGRSG---KOPTG----- 375
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 374 LRQHRGLMCNONGAGLGOPQASPTDFPNAHLDAYVWIKRPGESDGTSAADPTTGKSD 433
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 376 -QKAQDWCNAKGTGFLRSTNTGD---ALADAFAVWVKRGESDGS-----DTSARKYD 427
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 434 PWCDDPTTYSYGLVNALPNSPIAGQWFAPODQVLVANARPA 475
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 428 YHCG-----LDALKPAPAGTWFQAYFQQLDNNAPS 459
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
US-09-286-691-25 : Sequence 25, Application US/09286691
: Patent No. 6190189
: GENERAL INFORMATION:
: APPLICANT: Li, Xin-Liang
: APPLICANT: Ljungdahl, Lars G.
: APPLICANT: Chen, Huizhong
: TITLE OF INVENTION: Cellulases and Coding Sequences
: FILE REFERENCE: 42-96
: CURRENT APPLICATION NUMBER: US/09/286,691
```

```

: CURRENT FILING DATE: 1999-04-05
: EARLIER APPLICATION NUMBER: US 60/027,883
: EARLIER FILING DATE: 1996-10-04
: EARLIER APPLICATION NUMBER: PCT US97/18008
: EARLIER FILING DATE: 1997-10-03
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 25
: LENGTH: 360
: TYPE: PRT
: ORGANISM: Fusarium oxysporum
US-09-286-691-25

Query Match
Best local Similarity 8.5%; Score 555; DB 4; Length 360;
Matches 142; Conservative 64; Mismatches 147; Indels 68; Gaps 14;

QY 58 DNPYAGATFEPVNPYMAOEVOSEAA---NOTNATLAKMRVSTYSTAVWMRIAINGVN 114
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 DNPYAGVDLMANNYRSEVNNLAIVPKLSGAKATAAKADVPSFO---WMDTYDHSILME 59
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 115 GCGGLTYYIDAALSOQOQTPEVIEIYIDLPGRCALASNGELPATAGTQYETQYI 174
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 -----DTLDIRKANRAG-GKYAGQFVYVDLPNRDCAALASNGEYSLDKGANKRYA-YI 112
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 175 DPASISLNPKYSRLIYIIEPDSLPAVNTMSIOACATAVRYEOGIEVALTKLHAIP 234
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 113 AKIKIGLON--YSDTYVYIIEPDSLANTVTLNVDKCAKASAKELVYAIKELN-LP 169
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 235 NYIYMDAHSGLGMPNNSGYVOEKVLAASIGVNGIDCFVTNTANTYPLKEPMTA 294
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 170 NVSMYLDAGHGMVGWNPANGPAKLYAQIYKDAKPSRVRGLVTNVSNN----- 220
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 295 TQOVGQPVESANFYQWNPDIIDEADYAVDLYSRVLAAGPSSIGMLIDFLRMGNGPNEP 354
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 221 -----GWLKSTKPDYTESNPYDEQRY-INAFAFLAQEGSWNVKFLVDGRSG-----KOP 271
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 355 TGPSTATDVNTFVNSKIDLRQHRGLMCNONGAGLGOPQASPTDFPNAHLDAYVWIKPP 414
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 272 TG-----QKAQDWCNAKGTGFLRSTNTGD---ALADAFAVWVKPG 310
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 415 GESDGTSAADPTTGKSDMPCPTTYSYGLVNALPNSPIAGQWFAPODQVLVANARP 474
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 311 GESDGT-----DTSARKYDHCG-----LDALKPAPAGTWFQAYFQQLDNNANP 357
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 475 A 475
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 358 S 358

RESULT 11
US-09-687-147-25 : Sequence 25, Application US/09687147
: Patent No. 6268198
: GENERAL INFORMATION:
: APPLICANT: Li, Xin-Liang
: APPLICANT: Ljungdahl, Lars G.
: APPLICANT: Chen, Huizhong
: TITLE OF INVENTION: Cellulases and Coding Sequences
: FILE REFERENCE: 42-96a
: CURRENT APPLICATION NUMBER: US/09/687,147
: CURRENT FILING DATE: 2000-10-12
: PRIOR APPLICATION NUMBER: US 60/027,883
: PRIOR FILING DATE: 1996-10-04
: PRIOR APPLICATION NUMBER: PCT US97/18008
: PRIOR FILING DATE: 1997-10-03
: PRIOR APPLICATION NUMBER: 09/286,691
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 25
: LENGTH: 360
```


Db 405 NTGD---SLDSFVWVKRGECDDGTSDSSAP-----RFDSHC-----ALPDALQAP 448

QY 456 IAGWFPRAQFQOLYANARPA 475

Db 449 QAGAMFOAYFVOLLNANPS 468

RESULT 14

US-08-276-213-3

Sequence 3, Application US/08276213

Patent No. 5536655

GENERAL INFORMATION:

APPLICANT: Thomas, Steven

APPLICANT: Laymon, Robert

APPLICANT: Himmel, Michael

TITLE OF INVENTION: GENE ENCODING FOR THE EI ENDOGLUCANASE

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: National Renewable Energy Laboratory

STREET: 1617 Cole Boulevard

CITY: Golden

STATE: CO

COUNTRY: USA

ZIP: 80401-3393

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/276,213

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: O'Connor, Edna

REGISTRATION NUMBER: 29,252

REFERENCE/DOCKET NUMBER: NREL IR# 94-08

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303)231-1000

TELEFAX: (303)231-1098

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 521 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: N-terminal

US-08-276-213-3

Query Match

Best Local Similarity 8.3%; Score 543; DB 1; Length 521;

Matches 138; Conservative 48; Mismatches 104; Indels 100; Gaps 16;

QY 869 GVTSGQDEHYVQINNEW-----NSSAQCLITINAT-----GATVYSTANFSGGEGGADAT 919

Db 202 GVOYNGSY-----WGGNLOGAGQYVVLNPNRLVYSAHDVATSYVP-QTFESDPT 254

QY 920 YPSIYKGC---HWGNCITKNYGMPIQISQSAVYSMTTQVSSSAVVAVDIWNSTPT 976

Db 255 FPNMNPGLMKNKMGYLENQNIA-PVWLGEFGTLLQS----- 289

QY 977 TTGQPNGETEIMILNSRGVOPFGSQTATGYTAGHTWNVWQGOOTSWKIISYLLPGAT 1036

Db 290 TTDD-----TWLKL--VQYLRPTAQYAD--SFQTFW-----SWN----- 322

QY 1037 SISLNDLKAIRADAARGLSTFSD-----YLDVEAGFEIWMGGGIGSNSFSVS 1086

Db 323 -----PDSGDTGILKDDMQVTVKDGYLAPIKSIIDPVGASASPSQSPSPS 371

QY 1087 VT-----SGTSSPTPSPSPPTPSPPT--TPSPSPSPSPSTSSP-SSGVACRATYYVN 1138

Db 372 VSPSPSPSPSASRPTPTPTPTPTASPTPLTPTATPTPTASPTPTASGACRATSYOVN 431

QY 1139 SDMSGFTATVYVNTGSRATNGWTAVMSPGNGQVTVYVWMTALQSGASVATATLSTYNN 1198

Db 432 SDMGNGFTVYAVNTSGSVATKRTVTSWTFPGNGQITITSMNAATYQNGQSTYARMMSYNN 491

QY 1199 VIQPGQSTTFPGNGSYSGTNAAPTLSTCTAS 1228

Db 492 VIQPGQNTTFGFOASVYTGSMNAAPTVACAAS 521

RESULT 15

US-09-142-759-1

Sequence 1, Application US/09142759A

Patent No. 6127160

GENERAL INFORMATION:

APPLICANT: YAMANOE, Takashi

APPLICANT: MATANABE, Manabu

APPLICANT: HAKAYA, Toru

APPLICANT: SUMIDA, Naomi

APPLICANT: AOYAGI, Kaoru

APPLICANT: MURAKAMI, Takeshi

TITLE OF INVENTION: PROTEIN HAVING CELLULASE ACTIVITY AND PROCESS FOR

PREPARATION OF CELLULOSE

FILE REFERENCE: 051673

CURRENT APPLICATION NUMBER: US/09/142,759A

CURRENT FILING DATE: 1998-09-14

EARLIER APPLICATION NUMBER: PCT/JP97/00824

EARLIER FILING DATE: 1997-03-14

EARLIER APPLICATION NUMBER: P. HET-8-84479

EARLIER FILING DATE: 1996-03-14

NUMBER OF SEQ ID NOS: 30

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 1

LENGTH: 457

TYPE: PRP

ORGANISM: Acremonium cellulo]lyticus

US-09-142-759-1

Query Match

Best Local Similarity 8.3%; Score 540; DB 3; Length 457;

Matches 150; Conservative 60; Mismatches 171; Indels 82; Gaps 16;

QY 16 GTTRMPAISKRIRAGVLAGAVSIASIVPLAMQHPAIAATHVDNRYAGATFEVNPYAOE 75

Db 74 GTT-SPTTTTAKSTATTAAASG-----NFSQYOLANPYISE 115

QY 76 VOSEANOTNATLAAKRVSTYSTAVVMDRIAALINGVNGRGLTYYL-DAALSOQOQTT 134

Db 116 VHTLAIPLSLTGLAAATAKAEIPFVWLDTAAKV-----PTMGTYLANIEAAKACAS 169

QY 135 PEVIRI-IYVILPGRDCAALNSGELPATNAGIQYETROYIDPLASISNKRYSIRVY 193

Db 170 PPIAGIFVYVLDPRDCAALASNGEYTVANNGVAKA-YIDSYAQLK--AYDPVHTIL 226

QY 194 IIEPDLNNAVYTNMSIOACATAVPYEEOGIEVALKLAIIPVYLYMNAASGMLGNPN 253

Db 227 IIEPDLNNAVYTNMSIOACATAVPYEEOGIEVALKLAIIPVYLYMNAASGMLGNPN 285

QY 254 ASGYQEVQKVLNMSIGVNGIDGFEVNTANTYPLKEPMTATQOYGOPVESANFYQNP 313

Db 286 LSPAQLFATYVKNASAPASIRGLATNVAHYN-----AMSSISP-----SYTSGDS 332

QY 314 DIDEADVAVDLXSRVYAGFPSSIGMLDITLRNGGSGNEPFGSTADVTFTVQSKID 373

Db 333 NYDEKLYINALSPILTSNGWPA-HFIMDTSRNG---VQPT----- 369

QY 374 LROHRLGNCNONGAGLGPPOASPTDFPNNAHLDVAVVYKPGESDGTSAASDPTTGKSSD 433

Db 370 KOQAMGDMCNVIGTGFVQPTTNTGD-----PLEDAFVWVKPGESDGTSSNS---ATRYD 422

Oy 434 PMCDPTTSTYGVLTNALPNSPIAGQWFPQFDOLVANARPAV 476
Db 423 FHCG-----YSDALQPAPEACTWFFQAYFVOLLTNANPAL 456

Search completed: August 29, 2002, 16:22:32
Job time: 362 sec

THIS PAGE BLANK (USPTO)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)